PANCREATIC CANCER IMMUNOTHERAPY: T-CELL RECEPTOR SEQUENCING AND PERSONALIZED GENE FUSION VACCINES

by

Alexander Hopkins

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
August, 2017

Abstract

Immunotherapy has struggled to reduce the high mortality rate of pancreatic cancer, in part due to the low mutation burden and immunosuppressive microenvironment associated with the disease. Early successes have been limited to a small number of patients, highlighting the difficulty of attempts to modulate immune systems that have evolved to be unique in each patient.

To better understand the mechanisms of anti-tumor immune responses and elucidate the differences between clinical responders and non-responders, an analysis of T-cell receptor repertoires of 106 pancreatic cancer immunotherapy patients was performed using a newly developed analysis package, `immunoSeqR`. The data provides new insights into the effects of several immunotherapies used in current clinical trials, including immune checkpoint blockade. Additionally, diverse baseline repertoires and treatment induced clonal expansion were found to be significantly associated with clinical response.

Cancer vaccines must provide a source of antigens to direct an immune response, and while neoantigens arising from point mutations are often utilized for this purpose, fusion antigens, derived from genomic rearrangements are frequently overlooked. Seventeen putative fusion neoantigens were identified in two pre-clinical models and tested for their ability to induce immune responses and protect against tumor challenge when administered as a peptide vaccine. Antigen identification from next generation sequencing data was identified as a critical bottleneck for the development of personalized fusion vaccines.

Analysis of the T-cell receptor repertoire in clinical trial patients provides new insights into the effects that vaccines, radiation, chemotherapy and checkpoint blockade have on the complex lymphocyte compartment. Current vaccine approaches typically proceed without knowledge of whether antigens included in the vaccine are shared by the patient. Personalized vaccine approaches solve this problem, and advances in technology are reducing the cost and time required for their development. Together, these two approaches represent significant steps in closing the response gap in pancreatic cancer immunotherapy.

Thesis Committee

Elizabeth Jaffee, MD (Primary Advisor)
The Dana and Albert "Cubby" Broccoli Professor of Oncology
Co-Director of the Gastrointestinal Cancers Program
Associate Director for Translational Research
The Sidney Kimmel Cancer Center at Johns Hopkins

Victor Velculescu MD, PhD (Chair)
Co-Director of Cancer Biology
Professor of Oncology
The Sidney Kimmel Cancer Center at Johns Hopkins

Eric Lutz, PhD
Senior Principal Scientist
WindMIL Therapeutics

Robert Scharpf, PhD
Assistant Professor
Department of Oncology
Assistant Professor
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

Acknowledgments

In the more than 11 years that have elapsed since I first picked up a pipette, I have benefited from the mentorship of many great scientists. Dr. John Goudie and Mr. Michael Sinclair introduced me to science and are fantastic educators. Dr. Michael Laub and Dr. Josh Modell guided me through my first years as a scientist and reminded me always that science should not only be rewarding but also fun. Dr. Todd Golub and Dr. David Thomas gave me the confidence to trust myself as an independent scientist at one of the most prestigious institutions in the world. Finally, Dr. Eric Lutz and Dr. Elizabeth Jaffee guided me through the difficult and rewarding process of earning this degree.

To my parents, John and Peggy: when a doctor and a nurse show their children videos of brain surgeries during dinner, two of them will find it disgusting and the other will become a scientist. I can not overstate the impact that your compassion, dedication and patience has had on my life. My sisters improve the lives of everyone around them, mine most of all. Last, Dr. Chelsea Black, you continue to improve my life as a scientist and a person, and I hope you never stop.

I can not thank you all enough, and getting here would have been impossible without you.

Table of Contents

List of Tables

List of Figures

Chapter 1

Background

1.1 Pancreatic Cancer

Pancreatic cancer consistently ranks among the most deadly cancer types. In the next year, an estimated 53,000 people will be diagnosed in the United States, and only around 7% will survive for 5 years[1]. Around 20% of cases are attributed to smoking, and around 10% of patients have a family history of the disease[2]. By far the most abundant subtype is pancreatic adenocarcinoma, which typically arises in the epithelial cells of the pancreatic ducts.

Ideally, treatment begins with surgery, however 80% of patients have unresectable disease at diagnosis[3]. Fluorouracil or gemcitabine based chemotherapy and radiation remain the front-line therapy, however, a wide variety of immunotherapies are emerging as promising new treatment avenues.

1.2 Cancer Immunotherapy

The first recorded use of a therapy which harnessed the immune system to combat tumors was in 1891, when a physician in New York injected bacteria into tumors after noticing spontaneous remissions in sarcoma patients with concurrent infections. A century later, researchers demonstrated that cytotoxic T-lymphocyte associated protein 4 (CTLA4), expressed on the surface of T-cells, acts to prevent T-cell activation

by antigen presenting cells[4]. Shortly thereafter, the same group demonstrated that inhibition of this molecule using a therapeutic antibody allows the immune system to target tumor cells in mice[5]. This observation was quickly confirmed in humans when anti-CTLA4 became the first therapy to extend the lifespan of advanced melanoma patients[6]. This re-discovery of the possibilities of cancer immunotherapy ignited the current field of research, which has further explored the potential of antibodies targeting immune checkpoint inhibitors, returned to the century old method of intratumoral injection of bacteria, and branched into countless other areas of investigation.

One such branch of investigation includes cancer vaccines, which can either be developed for prevention or treatment of existing cancers. Allogeneic pancreatic cancer vaccines, genetically engineered to secrete granulocyte macrophage colony-stimulating factor (GM-CSF), termed GVAX, were first introduced as a potential pancreatic cancer therapy in 2001[7]. Phase I trials showed promising immune infiltration, and Phase II trials indicated a slight clinical benefit when compared to chemotherapy[8]. Like GVAX monotherapy, trials of anti-CTLA4 alone in PDA were underwhelming[9], which is why current clinical trials pair the vaccine with the checkpoint inhibitors anti-CTLA4[10] and anti-PD1[11], as well as genetically modified, tumor antigen-expressing *Listeria monocytogenes*[12]. The goal of these combinations is to simultaneously provide a broad range of potential tumor-specific antigens and remove the barriers to T cell activation in the periphery (with anti-CTLA4) or the tumor (with anti-PD1).

Clinically, GVAX is administered as a mixture of two human pancreatic cancer cell lines, modified to secrete GM-CSF. Because it is allogeneic, altered peptides resulting from patient-specific mutations (neoantigens) are unable to be targeted, and immunogenicity is contingent on common mutations that occur in the cell lines and the patient's tumor (cancer antigens). A variety of novel approaches are attempting truly personalized cancer vaccines, in which each patient's tumor DNA is sequenced

and potential neoantigens are identified, synthesized, and administered to the patients as a therapeutic vaccine via a number of possible methods. This method has the potential to target as many good quality tumor antigens as possible, but is often prohibitively expensive and frequently requires more time to develop than is possible for patients with advanced disease.

1.3 Immune Repertoire Profiling

Historically, descriptions of a patient's immune health relied on laborious and relatively unsophisticated laboratory methods, including blood counts. While clinically instrumental in a number of disease cases, these methods fail to identify subtle changes in the population of immune cells which may be earlier and more accurate markers of disease. Researchers interested in examining the repertoire of B or T lymphocytes had a limited set of tools available. Typically, staining of T lymphocytes with fluorescent labeled multimers of major histocompatibility complex (MHC) molecules provided the broadest possible look at the repertoire, but this powerful method suffers from several key limitations. First, T-cells see their cognate antigen in the context of specific MHC molecules. In the population, hundreds of MHC allele variants exist, and in an individual, several different alleles can be expressed. This allele restriction problem means that MHC multimer staining can only look at one allele at a time. Additionally, antigenic peptides must be known *a priori* in order to be tested in this assay, so screening efforts of large libraries are necessary to identify novel antigens.

A complementary approach involves amplification of specific genomic rearrangements associated with lymphocyte development. Somatic recombination is a unique feature of the adaptive immune system which involves the removal and rearrangement of individual gene segments inside each lymphocyte, followed by the insertion of random non-templated nucleotides. The result of this process is that each newly generated immune cell contains a unique receptor with complementary determining

regions (CDRs) created by both the combinatorial mechanism (three gene segments V, D and J are combined, each drawn from a large pool of possible locii), and random addition and deletion of nucleotides. The human body contains around 25-100 million unique T-cell clones, comprising an estimated 10^{12} total T-cells[13], but the process of VDJ recombination allows for an estimated 10^{16} possible unique T-cell receptors[14]. While PCR amplification of the CDR regions has been possible for decades, it took the development of modern parallel sequencing technologies for the quantification of the repertoire to become possible.

The foundations for the current technology were laid in 2009, when it was demonstrated that sequencing of the CDR3 region of the TCRβ chain allowed rapid quantification of the TCR repertoire and tracking of individual clones over time[14]. The throughput of this assay, as well as its depth make it a powerful tool for immune monitoring, however there are notable drawbacks. First, unlike MHC multimer staining, high throughput T-cell receptor Vβ sequencing (HTTCS) does not allow identification of the cognate antigen associated with each sequence. In fact, because the technology sequences only the β chain, and peptide recognition requires both the α and β chains, antigen identification directly from this data is impossible. Second, the technology can not itself distinguish between T-cell subtypes, whereas this would be easily accomplished in MHC multimer staining using multi-channel flow cytometry. The result is that if the subset population (i.e. $CD4^+$ vs $CD8^+$) is of interest, the populations must be separated and run as separate samples in the HTTCS assay.

Despite these challenges, this technology found early use as a clinical assay to monitor response to therapy in patients with lymphoblastic leukemia[15]. It was quickly adopted for research into infectious diseases[16] and eventually cancer immunotherapy, where current efforts focus on determining the mechanisms of new therapies, and predicting clinical responses by monitoring immune health.

Chapter 2

T Cell Receptor Sequencing in Pancreatic Cancer Immunotherapy Trials

2.1 Introduction

2.1.1 T Cell Receptor Profiling

The development of high throughput T cell receptor Vβ sequencing (HTTCS) has allowed the identification and temporal monitoring of clones with much greater sensitivity than previous methods. The technology uses a pool of 45 primers specific to all known variants of the V segment of the TCRβ gene, as well as 13 primers specific to known Jβ regions to amplify all rearranged TCR sequences in a given DNA sample. After amplification, the product is sequenced using Illumina sequencing technologies, and unique TCR sequences can be counted. The quantities of each primer are calibrated to normalize any variation in amplification efficiencies, and a computational method using control samples is used to further reduce noise. The resulting data set contains the nucleotide sequences of each unique T-cell receptor identified, along with an estimation of how many copies of that sequence were present in the sample.

Initial applications of this technology focused on quantifying minimal residual disease in patients with hematopoietic malignancies, in which a single malignant clone is tracked during the course of therapy[17]. It was quickly recognized, however,

that descriptions of the repertoire provided information about the overall health of the immune system.

HTTCS was used, for example, in immunotherapy trials in melanoma, where it revealed that inhibition of CTLA4 leads to an expansion of the T cell repertoire. This expansion was also correlated with increased toxicity[18]. Several additional groups have now correlated TCR repertoire metrics with various measures of clinical outcome[19,20,21], though the methods used vary widely between studies.

TCR repertoire studies of patients treated with anti-PD1 have focused primarily on the tumor repertoire, rather than the peripheral repertoire, due to the fact that anti-PD1 acts upon previously activated T-cells at the site of the tumor. Differences in peripheral repertoire metrics in patients treated with either anti-CLTA4 or anti-PD1 have been described[22], hinting that peripheral diversity is less important in the case of anti-PD1 treatment. Repertoire profiling in the tumor creates unique challenges for the technology and analysis. Formalin fixed, paraformaldehyde embedded tissues, for example, experience DNA degradation, resulting in poor sequencing quality. Additionally, low numbers of T-cells present in the sample (compared with peripheral blood) lead to estimates of diversity which do not accurately represent the larger population. Finally, T-cells do not distribute themselves randomly inside a tumor, but rather assemble into complex structures known as tertiary lymphoid aggregates[23]. Because of this non-random spatial distribution, the number and makeup of T-cells can vary greatly between slides cut from the same tumor block. Conflicting studies have been published showing a high degree of concordance between multiple samples taken from the same tumor in ovarian cancer[24] and very low concordance between multiple samples of the same renal cell carcinoma[25]. Despite these challenges, in at least one study, clinical responders have been shown to have a greater number of expanded clones, as well as increased repertoire clonality among tumor infiltrating lymphocytes[26].

One of the hallmarks of cancer immunotherapy has been surprising clinical responses in a subset of patients. Predicting which patients will respond to certain therapies has been elusive, due to the complexity of the immune response. HTTCS measures this complexity to an unprecedented degree, leading a number of groups to pursue repertoire metrics as possible biomarkers for immunotherapy. In metastatic breast cancer, for example, it has been demonstrated that a model using lymphopenia combined with repertoire diversity is predictive of clinical outcome[20]. Larger studies will be necessary to provide sufficiently large test datasets to determine the efficacy of these biomarkers, but HTTCS seems like a promising technology for this application.

In the present study, HTTCS is used to examine 343 peripheral and tumor TCR repertoires taken from 106 pancreatic cancer patients enrolled in 5 clinical trials of different immunotherapies. To analyze this number of TCR repertoire samples, an R package called `immunoSeqR` was developed, allowing efficient processing of count data and management of sample metadata. The results indicate that repertoire clonality and the number of expanded clones are both associated with clinical outcomes, but only in some therapeutic contexts.

2.1.2 Clinical Trials

Adjuvant and Neoadjuvant GVAX

Details of the J9988 clinical trial (NCT00084383, herein the Adjuvant trial) have been described previously[8]. Patients eligible for this trial had histologically diagnosed pancreatic adenocarcinoma which was successfully removed (R0 or R1) by surgical resection within 10 weeks of enrollment. Patients were excluded for a variety of criteria, including metastatic disease, autoimmune or allergic disease, recent corticosteroid use, and HIV infection.

Enrolled patients were given the first dose of GVAX (5×10^8 total cells) 8-10 weeks following surgery. Around 2 weeks after vaccination, patients received 5-fluorouracil and radiation therapy. Following chemoradiation, patients received up to 3 maintenance doses of GVAX at 4 week intervals, and additional boosts 6 months following maintenance (See Figure 2.1, top).

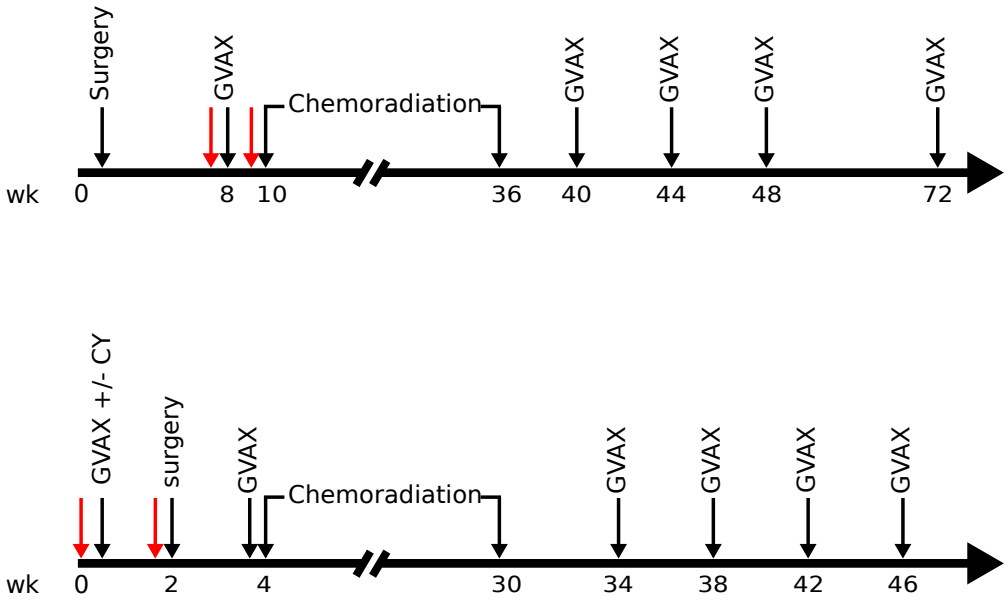

Figure 2.1: Timeline of the Adjuvant (top) and Neoadjuvant (bottom) GVAX studies. Black arrows indicate treatments and red arrows indicate blood draws taken during the study.

The neoadjuvant GVAX (J0810, NCT00727441) trial enrolled patients with resectable pancreatic adenocarcinoma, with no detectable metastasis[23]. Patients treated within 28 days with any anti-cancer therapy, steroid or immunosuppressive therapy, as well as patients previously treated with immunotherapy at any time were excluded from this study.

Patients enrolled in this trial were administered GVAX (alone or in combination with either oral or intravenous cyclophosphamide) two weeks before surgical resection. Following surgery, patients were treated with standard chemotherapy and radiation, followed by additional doses of GVAX every four weeks (see Figure 2.1, bottom).

Stereotactic Body Radiation with GVAX

Patients were enrolled in the J1179 (NCT01595321) clinical trial with documented pancreatic cancer within 10 weeks of successful surgical resection. Patients were excluded if they were given any prior therapy for pancreatic cancer, had metastatic disease, or any autoimmune or allergic disease.

Enrolled patients in arm 2 were treated with cyclophosphamide and GVAX two weeks before undergoing Stereotactic Body Radiation Therapy (SBRT). All patients (in arms 1 and 2) received SBRT, and 6 cycles of chemotherapy (FOLFIRINOX). Following chemotherapy, patients in arm 2 received additional maintenance doses of GVAX (see Figure 2.2).

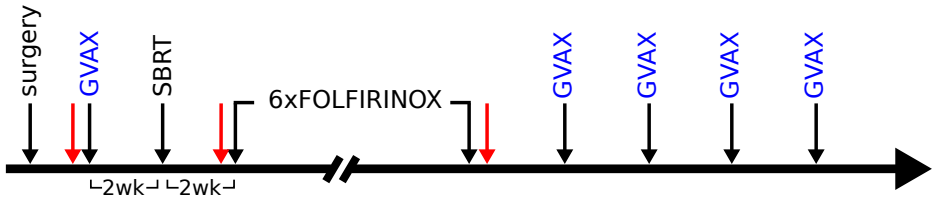

Figure 2.2: Timeline of the SBRT GVAX study. Black arrows indicate treatments, red arrows indicate blood draws taken during the study. All patients received treatments listed in black, and one arm received additional treatments listed in blue.

Anti-CTLA4 or anti-PD1 with GVAX

In the J0834 clinical trial (NCT00836407), patients were enrolled with local, advanced or metastatic pancreatic adenocarcinoma, previously treated with gemcitabine-based chemotherapy. Patients treated with anti-CTLA4 at any time, or any other cancer therapy within 28 days were excluded, as were patients with brain metastases or infection with HIV or hepatitis B or C.

Enrolled patients received 10mg/kg Ipilimumab only (arm 1) or 10mg/kg Ipilimumab + GVAX (arm 2) every three weeks for four cycles, followed by maintenance dosing (see Figure 2.3, top).

The J14113 clinical trial (NCT02243371) enrolled patients with metastatic pancreatic adenocarcinoma who failed one previous therapy for metastatic disease. Patients treated previously with anti-PD1, anti-PDL1/2, anti-CTLA4, GVAX, CRS-207 at any time or patients treated with steroids or immunosupressants within 28 days were excluded, as well as patients with HIV, hepatitis B or C infections.

Once enrolled, patients received two doses of GVAX followed by four doses of CRS-207, a strain of *Listeria monocytogenes* expressing mesothelin, a tumor antigen expressed by many tumors. Patients in arm 1 of this study were also administered 3mg/kg Nivolumab with all treatments (see Figure 2.3, bottom).

For both studies, baseline samples collected before the initial treatment (PRE) as well as after the third cycle (POST3) were used for analysis.

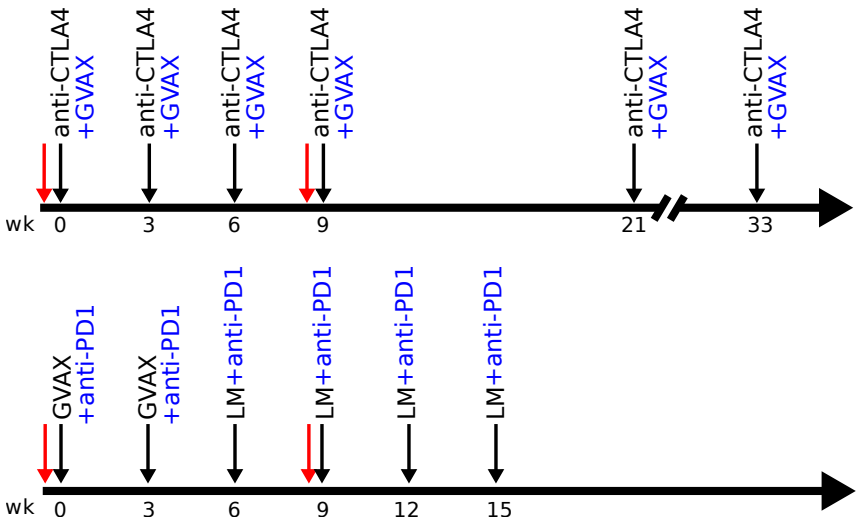

Figure 2.3: Timeline of the J0834 (top) and J14113 (bottom) clinical trials. Black arrows indicate treatments. Red arrows indicate blood draws used in this study. Black text indicates treatments received by patients in both arms, blue text indicates only patients in combination arm. LM: *Listeria monocytogenes*

2.2 Materials and Methods

2.2.1 Development of the `immunoSeqR` R Package

The `immunoSeqR` package for the R programming language was created to analyze data from large HTTCS experiments and organize experiment level metadata. The analysis of TCR data using this package involves the following steps:

1. Importing data from raw `.tsv` files
2. Merging samples together based on matching TCR sequences
3. Aggregating synonymous DNA sequences into single amino acid sequences
4. Calculating sample level and patient level statistics
5. Building a metadata dictionary
6. Graphical representation of data

The data is imported into a single `data.frame` containing all of the samples. Rows represent unique T cell clones and columns represent samples. When clones are not shared between samples, the table is completed by adding zeros. This results in a larger than necessary table, but saves time by not having to join samples at the time of analysis.

Samples are merged using the `iseqr_merge()` function:

```
all_files <- list.files("/path/to/tsv",pattern=".tsv")
# construct the dataset with iseqr_merge
ds <- iseqr_merge(all_files)
```

When sequencing from peripheral tissues, the data may contain TCRs with non-productive rearrangements, or stop codons within the CDR3 region. These are filtered out before proceeding.

Multiple synonymous nucleotide sequences can encode the same CDR3 amino acid sequence, so after the data has been joined, it must be aggregated, by summing the counts from synonymous sequences together. This is accomplished with the `iseqr_aggregate()` function:

```
ds_agg <- iseqr_aggregate(ds,inc_nt=FALSE)
```

which produces a data set containing individual unique CDR3 sequences in rows, and samples in columns. Summary statistics, including Richness and Clonality (discussed in Chapter 2.2.2) are calculated using the aggregated data, while expanded clones are calculated from the nucleotide level data.

Summary statistics can be calculated easily with the data in this format using the `apply()` family of functions. Clonality for example, can be easily calculated for all samples at once using

```
sapply(ds_agg,clonality)
```

Once the dataset is created, a metadata dictionary is created. This object is a `data.frame` in which each row corresponds to a column in the data set, and thus to a sample. The dictionary contains information about the sample such as which patient it was collected from, and at what time point. It also contains demographic information about each patient, including their age, sex, clinical response, Response Evaluation Criteria In Solid Tumors (RECIST) status, adverse event status, etc. After the metadata is collected, other statistics can be calculated, including the change in each metric before and after treatment, as well as the overlap between relevant samples. Finally, the dictionary and summary statistics are combined into a single `data.frame` (a `plot_ds` object), where each row contains the metadata and summary statistics for each sample.

The dictionary also enables the `immunoSeqR` package to more easily create plots with relevant layouts. By supplying the function `iseqr_plot_factor()` with the `plot_ds` object, a metric to plot, and a metadata field to compare, a plot (Figure 2.4) is created using the `ggplot` package, including 95% confidence intervals for the mean of each group.

```
iseqr_plot_factor(plot_ds,"Total Sequences","response")
```

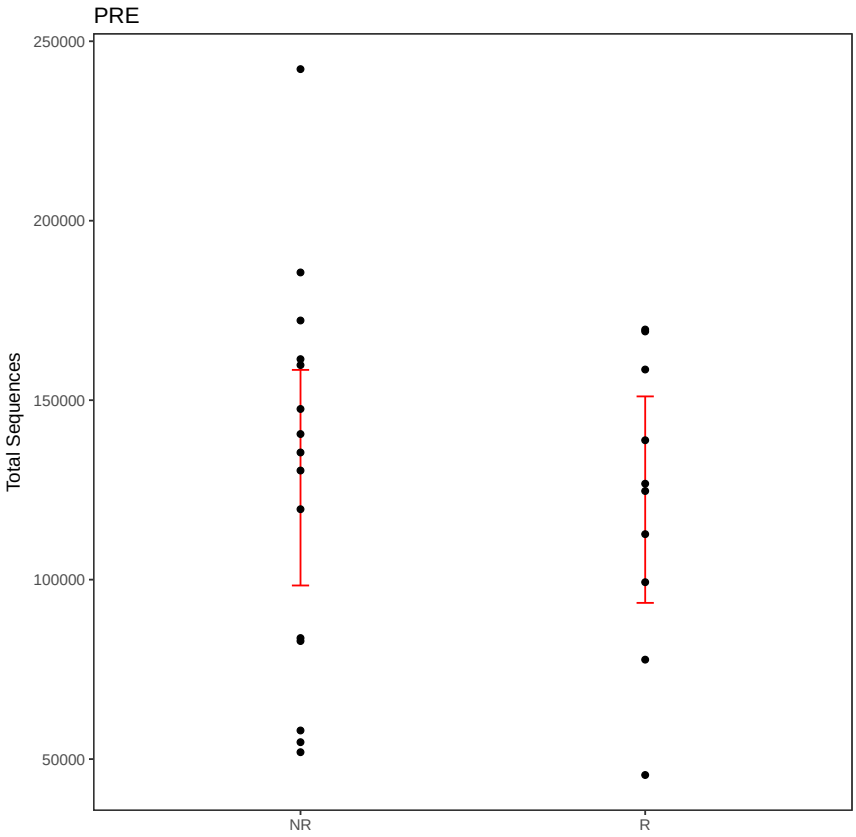

Figure 2.4: An example plot created with the `iseqr_plot_factor()` function. The metadata and summary statistics are stored in the `plot_ds` object, allowing the function to divide the data, calculate the 95% confidence interval (red bars) and plot the data.

2.2.2 Statistical Methods

Sample clonality is based on the normalized Shannon Entropy[27], and is computed as

$$C = 1 - \frac{-\sum p_i log_e(p_i)}{log_e(n)} \tag{2.1}$$

where p_i is the proportion of the ith clone, in a population with n clones. For paired pre and post-treatment samples, the normalized change in clonality was calculated as the base 2 logarithm of the ratio of the post-treatment clonality to the pre-treatment clonality.

The Morisita Overlap was calculated as

$$M = \frac{2\sum_{i=1}^{n} x_i y_i}{(D_x + D_y) n_x n_y} \tag{2.2}$$

where D_x represents the Simpson's Index (given below in equation 2.3) of sample x, n_x represents the Richness of sample x, and x_i represents the abundance of the ith clone in sample x (and the same for sample y). The Simpsons Index is calculated as

$$D = \sum_{i=1}^{n} p_i^2 \tag{2.3}$$

A Fisher Exact test was used to identify significantly expanded clones between pre and post treatment samples[16]. For each clone, the pre and post treatment counts are placed in the top row of a 2×2 contingency table, and the number of total T cells for each sample is placed in the bottom row. A one-sided Fisher's Exact Test is used to assign a p-value to each table, and thus each clone. To correct for multiple comparisons, the Benjamani & Hochberg Procedure[28] is used to maintain a False Discovery Rate of $\alpha = 0.05$. The number of significantly expanded clones for each patient is then calculated as the total number of clones with $p < 0.05$ after multiple comparison correction.

All metrics were compared across binary categorical sample groups using a non-parametric Wilcox rank-sum test. In cases where more than two groups were compared, an ANOVA was used to determine significance. For cases where metrics are compared to continuous variables (such as dose), a Pearson coefficient was used to assess the strength of correlation.

2.2.3 Preparation of Samples

Peripheral blood mononuclear cells (PBMC) were processed as previously described[29] and in accordance with the standard operating procedures of each clinical trial, and stored in liquid nitrogen until testing. PBMC samples were thawed, counted and frozen as a cell pellet, assuring that, for each patient, the same number of cells was sent for each time point. PBMC samples were sequenced using the immunoSeq assay[14] at the deep sequencing level.

Tumor samples from surgical resections and needle biopsies were formalin fixed and paraffin embedded in accordance with the standard operating procedures of the clinical trials. Slides from each tissue section were stained with Hemotoxylin and Eosin (H&E) and assessed for their tumor purity. For each patient, the tumor sections with the highest tumor content were selected and any sections containing intestinal tissue or other potentially confounding tissues were discarded. Ten 25μm sections were cut from these tissue selections for sequencing. Tumor sequencing was performed using the immunoSeq FFPE assay at the survey level.

2.3 Results

2.3.1 GVAX as Adjuvant and Neoadjuvant Therapy With Chemotherapy

Patients enrolled were treated with GVAX before chemoradiation, but either after (J9988, adjuvant) or before (J0810, neoadjuvant) surgical resection of the tumor (See Figure 2.1). At baseline, patients receiving neoadjuvant GVAX had significantly more clonal peripheral T cell repertoires than patients on the adjuvant trial, a trend which persisted after treatment (Figure 2.5).

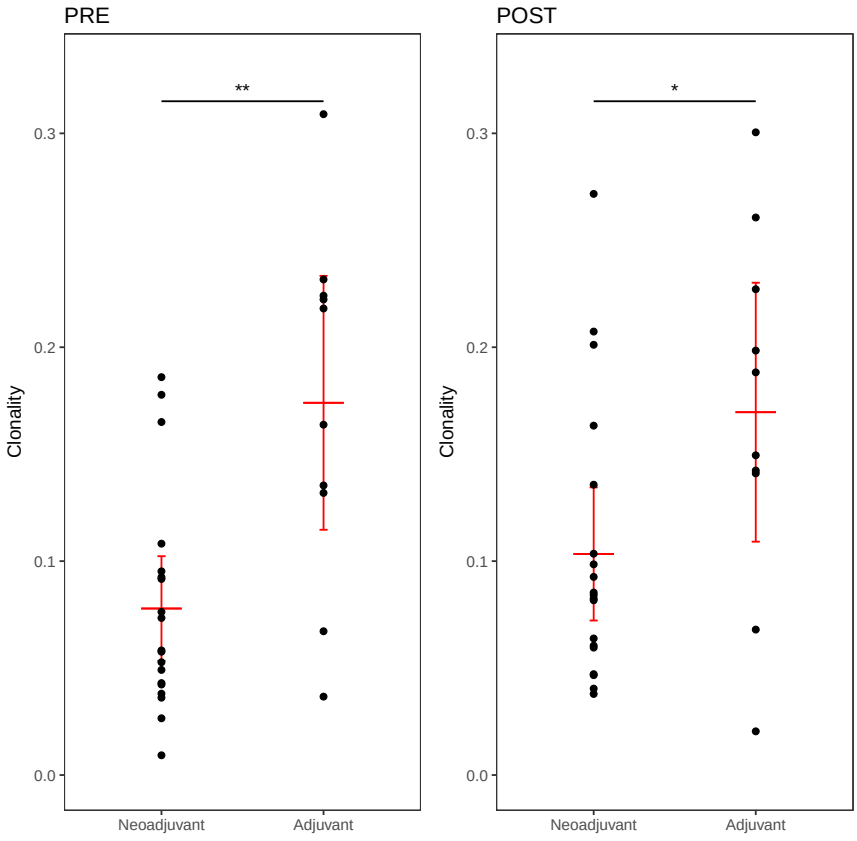

Figure 2.5: The clonality of all patients before (left) and after (right) treatment, separated by study.

To assess the effects of the different treatment modalities on individual patients, the fold change in clonality relative to baseline (log2 normalized) was used. The ma-

jority of patients on the neoadjuvant trial developed more clonal repertoires following treatment (increase in clonality, positive fold change) when compared to patients treated in the adjuvant setting (Figure 2.6).

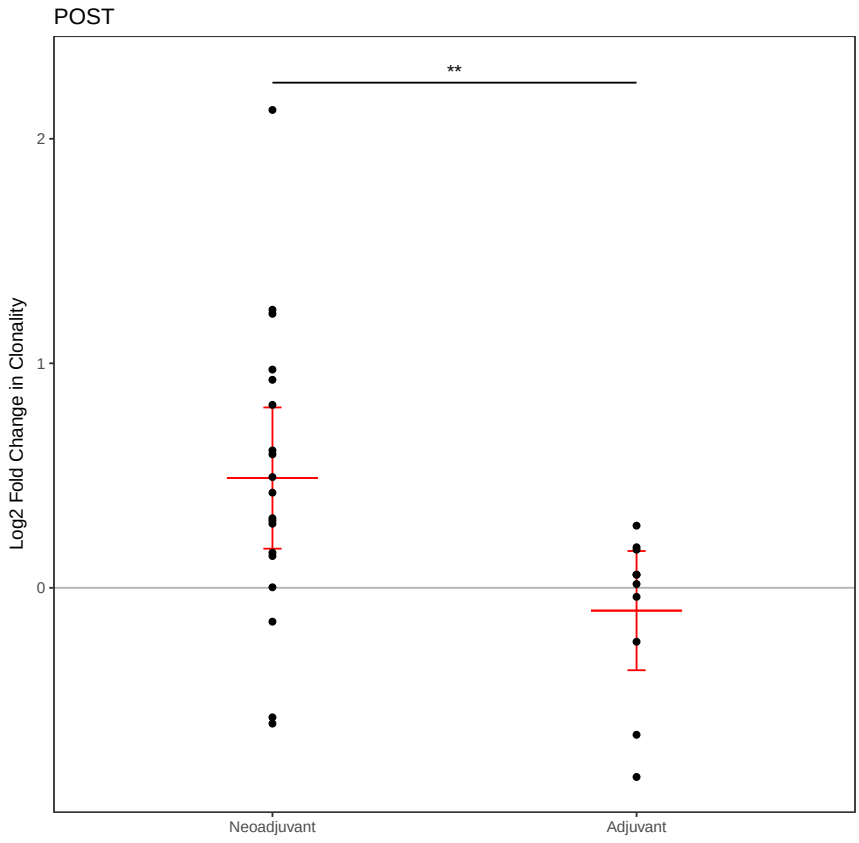

Figure 2.6: The Log2 normalized fold change in clonality from pre to post treatment of all patients, separated by study.

Patients treated in the neoadjuvant study experienced larger changes in their repertoire following treatment, indicated by significantly lower Morisita indices (a measure of similarity, Figure 2.8). No significant change between the trials was observed when the number of expanded clones induced in each patient was compared (Figure 2.7).

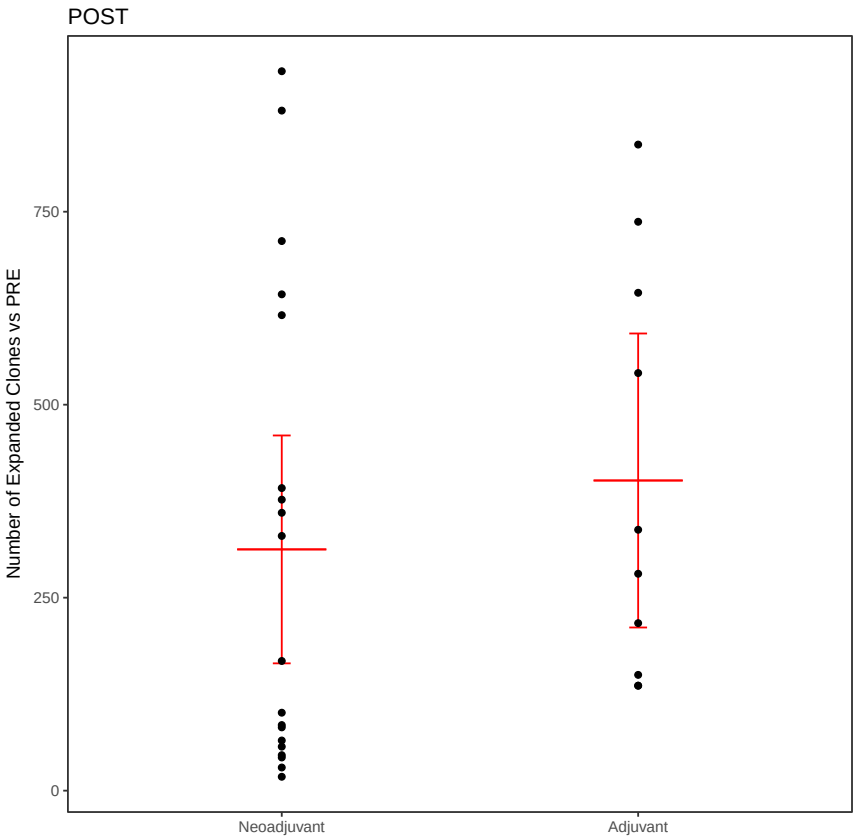

Figure 2.7: The number of significantly expanded clones following treatment in all patients, separated by study.

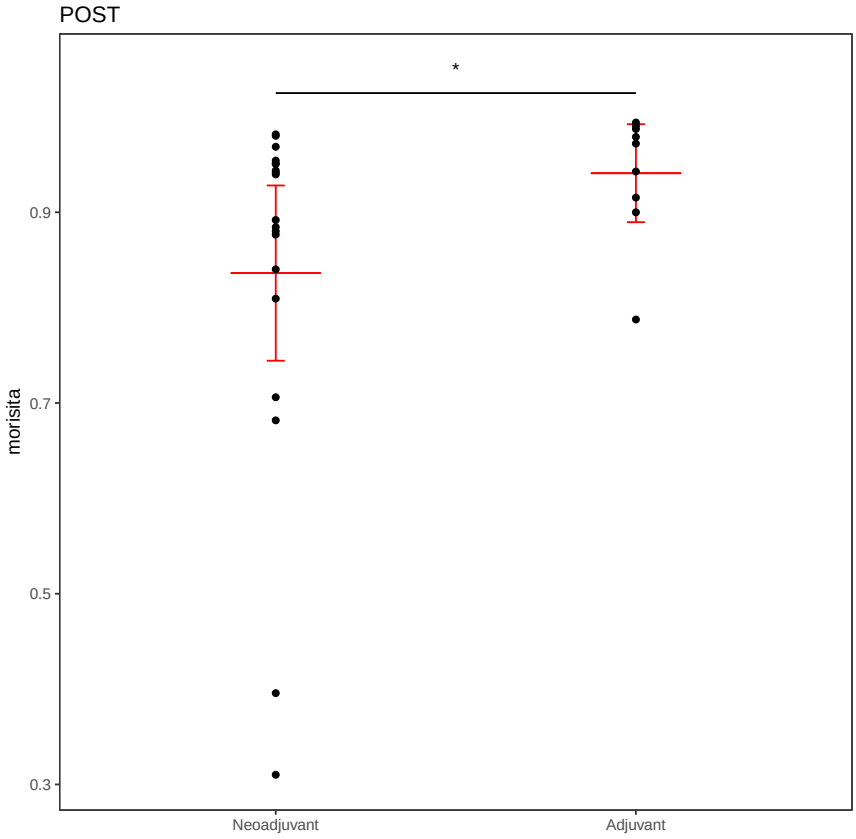

Figure 2.8: The Morisita index (a measure of similarity) of all patients separated by study

The relatively large variability of these metrics in the neoadjuvant study led to the hypothesis that the administration of cyclophosphamide (and its method of administration) may have a large effect on the T cell repertoire. To test this hypothesis, each treatment arm (including the adjuvant study for comparison) was assessed separately, using the same metrics as before. Significant group differences were identified using ANOVA. The average baseline clonality was significantly different among groups, but stabilized after treatment (Figure 2.9).

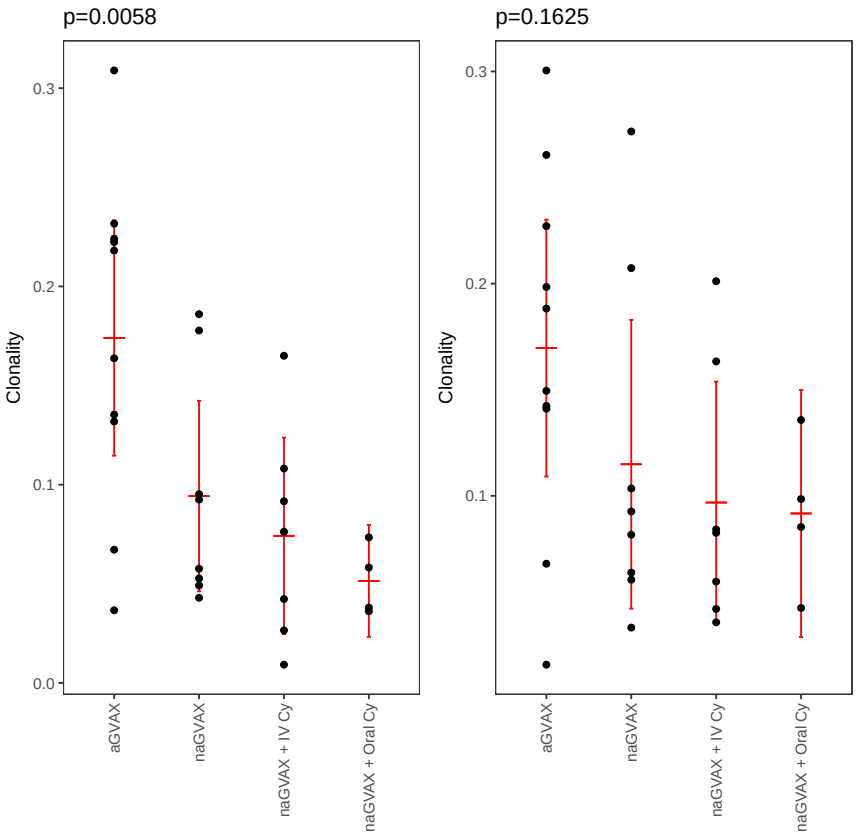

Figure 2.9: Pre-treatment (left) and post-treatment (right) clonality of each sample, separated by treatment arm. aGVAX represents adjuvant vaccination (vaccine is administered after surgery) and naGVAX represents neoadjuvant treatment (vaccine administered before surgery). P-values represent results of ANOVA analysis at each time point.

On a per patient basis, significant differences were observed between groups in the fold change of clonality relative to baseline (Figure 2.10). Patients who received oral or IV cyclophosphamide had larger increases in clonality following treatment than patients in the GVAX only arm.

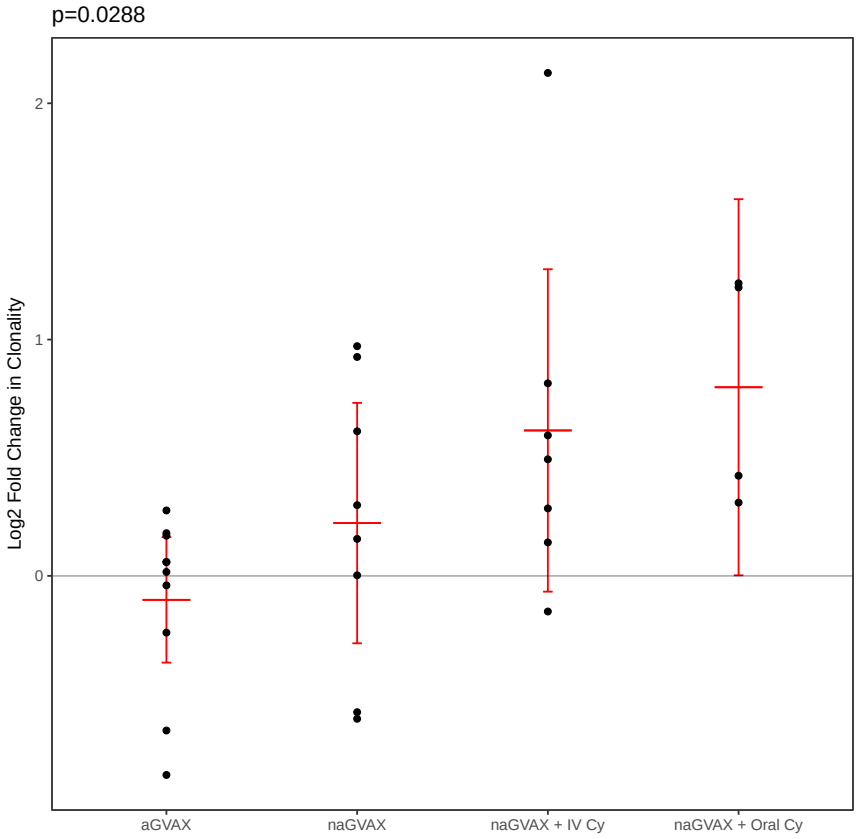

Figure 2.10: The Log2 normalized fold change in clonality from pre to post, separated by treatment arm. P-values represent results of ANOVA analysis.

The overall change in the repertoires (quantified by the Morisita index) was unchanged between treatment arms (Figure 2.12), as was the number of expanded clones (Figure 2.11).

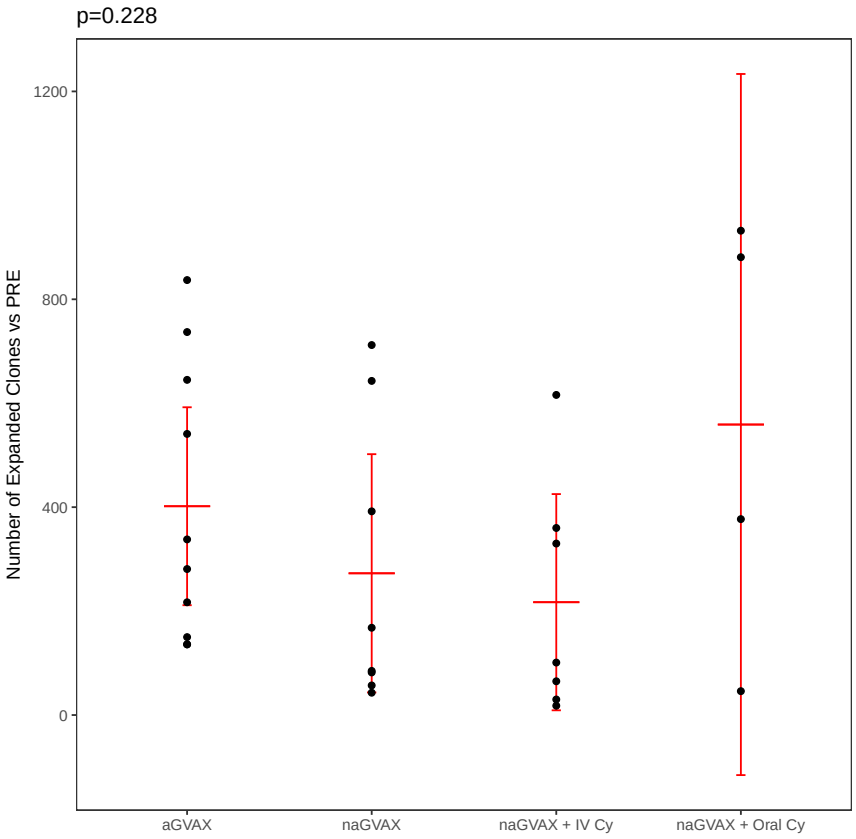

Figure 2.11: The number of significantly expanded clones following treatment in all patients, separated by treatment arm. P-values represent results of ANOVA analysis.

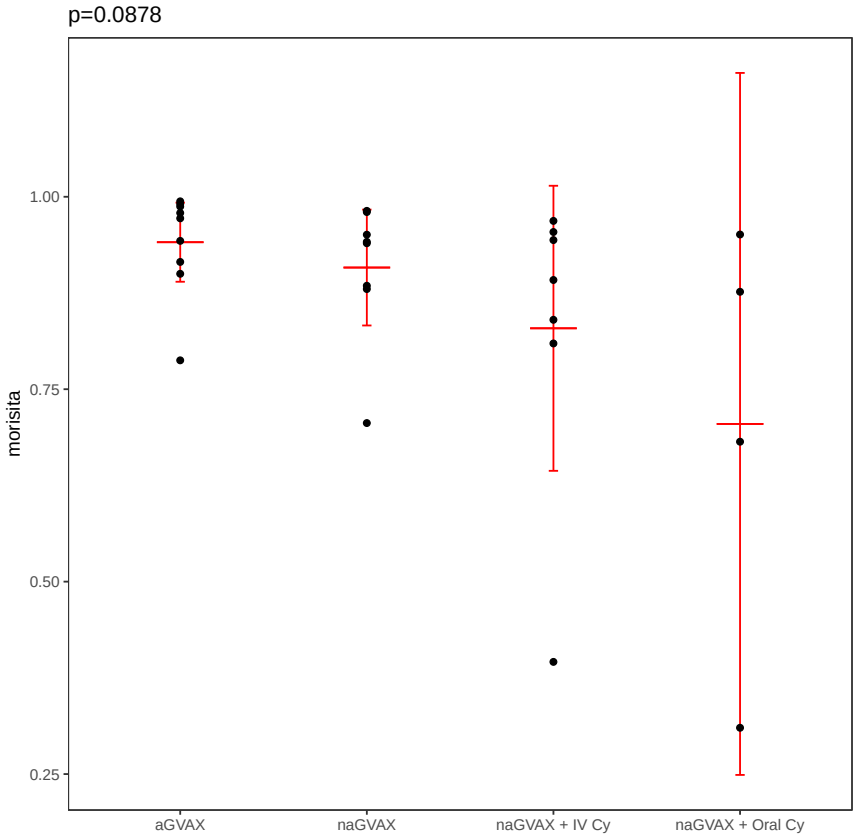

Figure 2.12: The Morisita index (a measure of similarity) of all patients of all patients separated by study, separated by treatment arm. P-values represent results of ANOVA analysis.

To better understand the mechanisms of successful anti-tumor vaccination, and to assess the possibility of HTTCS as a prognostic marker, patients were separated by clinical response (responders defined as >6 months overall survival) and assessed using the same series of metrics. In the adjuvant trial, clinical responders trended towards more diverse repertoires (before and after treatment), however this difference was not significant. No changes in clonality were observed in clinical responders in the neoadjuvant trial (Figure 2.13).

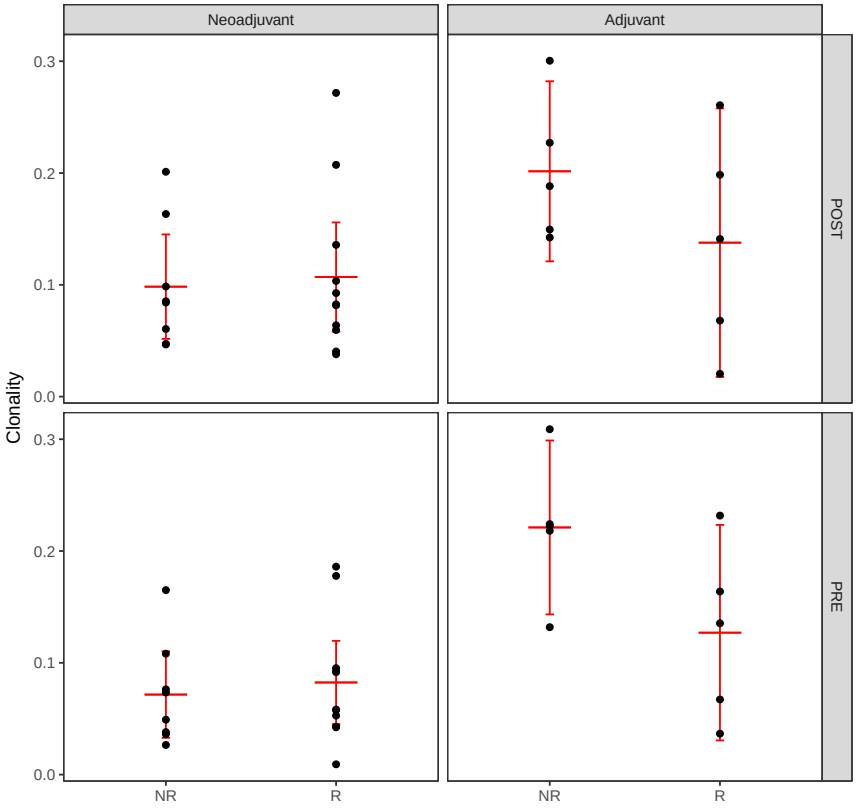

Figure 2.13: The clonality of all patients in the neoadjuvant (left column) and adjuvant (right column) before (bottom row) and after (top row) treatment, separated by clinical response (defined as at least 6 months overall survival).

As observed in the aggregate data, no difference was observed in the patient level fold change in clonality separating clinical responders from non-responders (Figure 2.13).

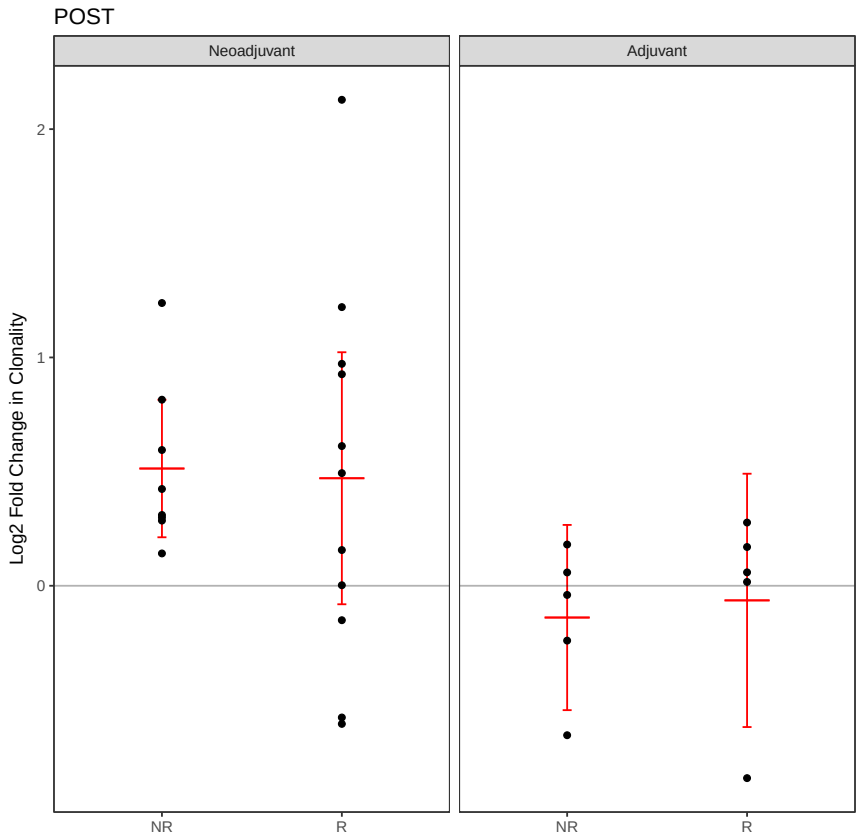

Figure 2.14: The Log2 normalized fold change in clonality from pre to post treatment, of patients on the neoadjuvant (left) or adjuvant (right) trials, separated by clinical response.

Clinical responders were likewise indistinguishable from non-responders using the Morisita index (Figure 2.16) or the number of expanded clones (Figure 2.15), though a counterintuitive trend was observed in the latter, with clinical responders having a slightly (and not significantly) lower number of expanded clones.

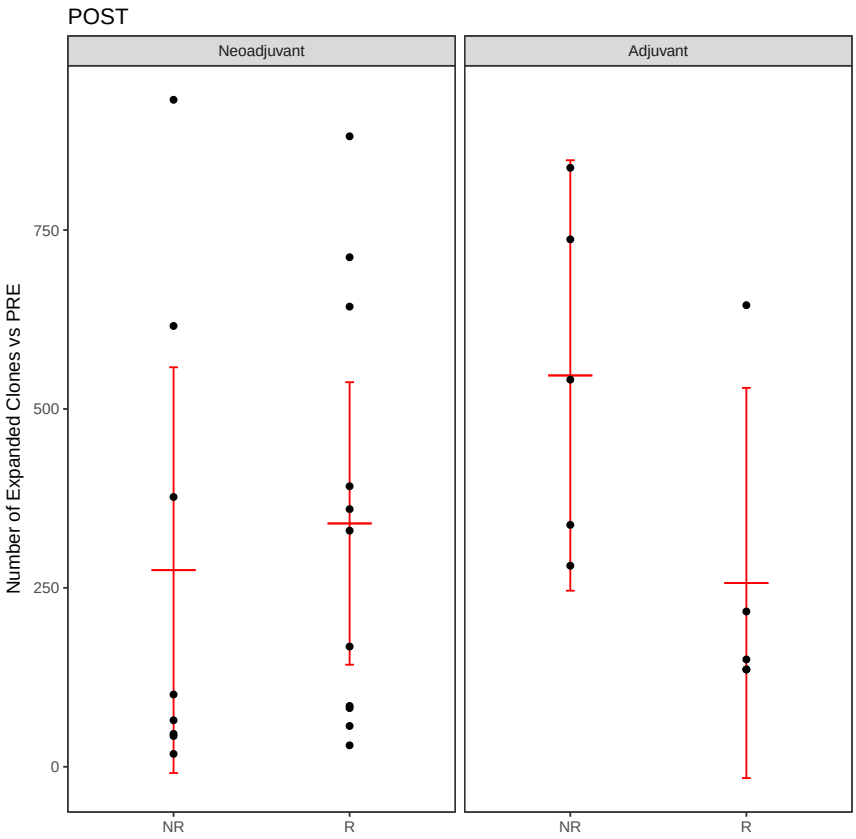

Figure 2.15: The number of significantly expanded clones after treatment in patients in the neoadjuvant (left) or adjuvant (right) trials, separated by clinical response.

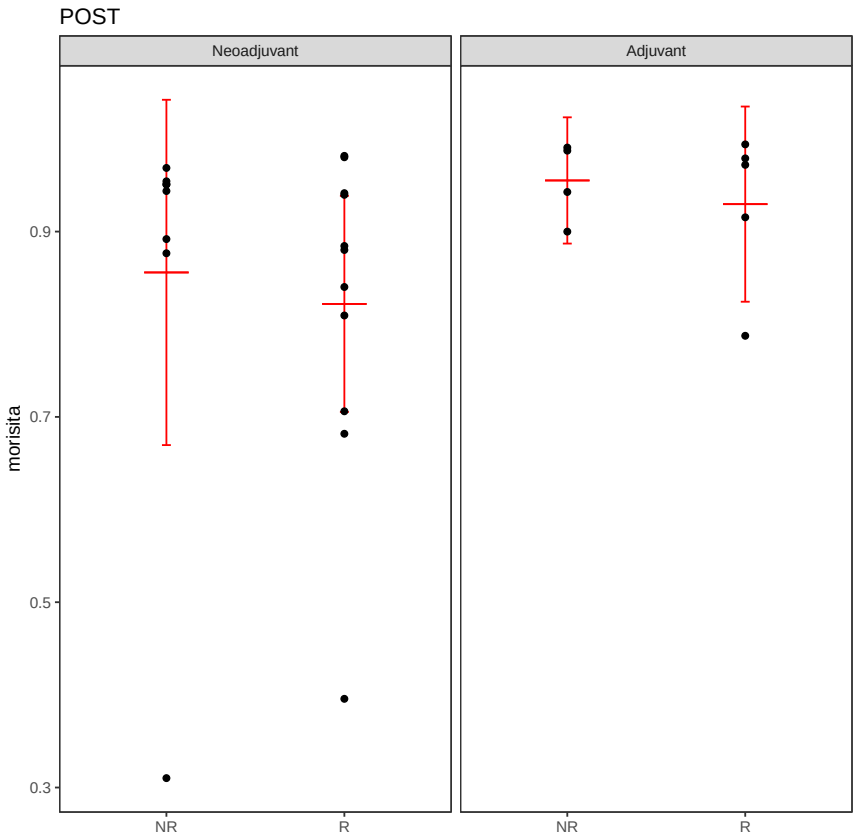

Figure 2.16: The Morisita index of all patients in the neoadjuvant (left) and adjuvant (right) trials, separated by clinical response.

For a subset of patients on both trials, material from the surgical resection was available for TCR repertoire analysis. Due to the comparatively low numbers of T-cells present in the tumor, it was first necessary to assess the number of total TCRs sequenced in each sample. The clonality metric is only a valid representation of the repertoire in cases where the number of T-cells sequenced is greater than 100. In tumor samples taken from patients in these studies, the vast majority were above this threshold (Figure 2.17). Any samples below the threshold were discarded for further analyses.

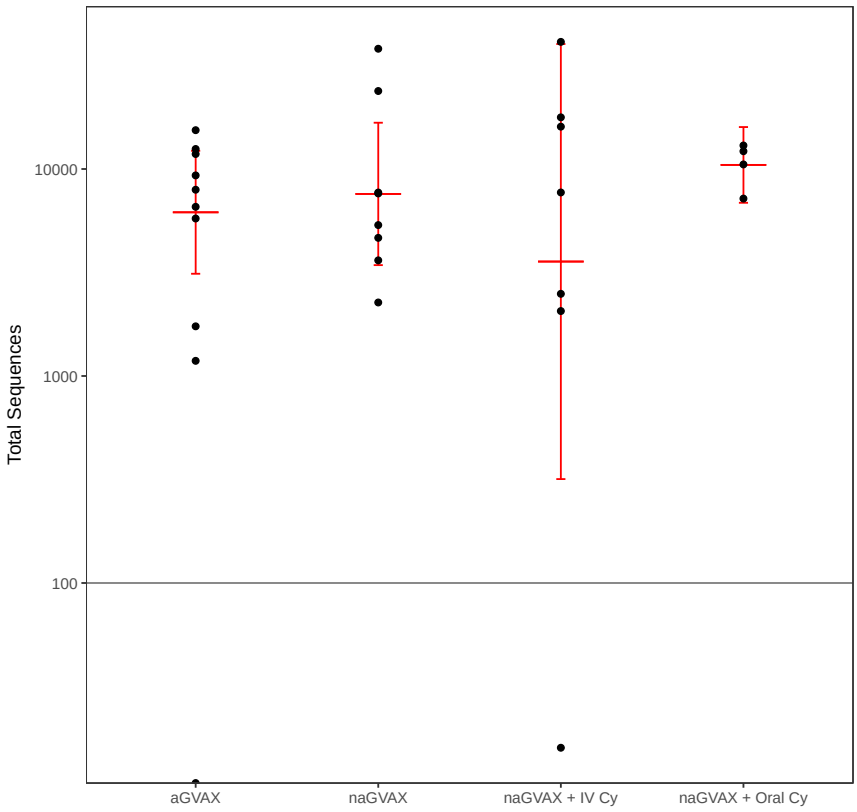

Figure 2.17: Total number of TCR sequences for tumor samples from patients in the adjuvant and neoadjuvant trials

No significant difference was observed between treatment arms (ANOVA, $p = 0.43$) or between clinical response groups in the clonality of tumor repertoires (Figure 2.18). Due to the lack of biopsy material available, pre and post-treatment comparisons were impossible in these trials.

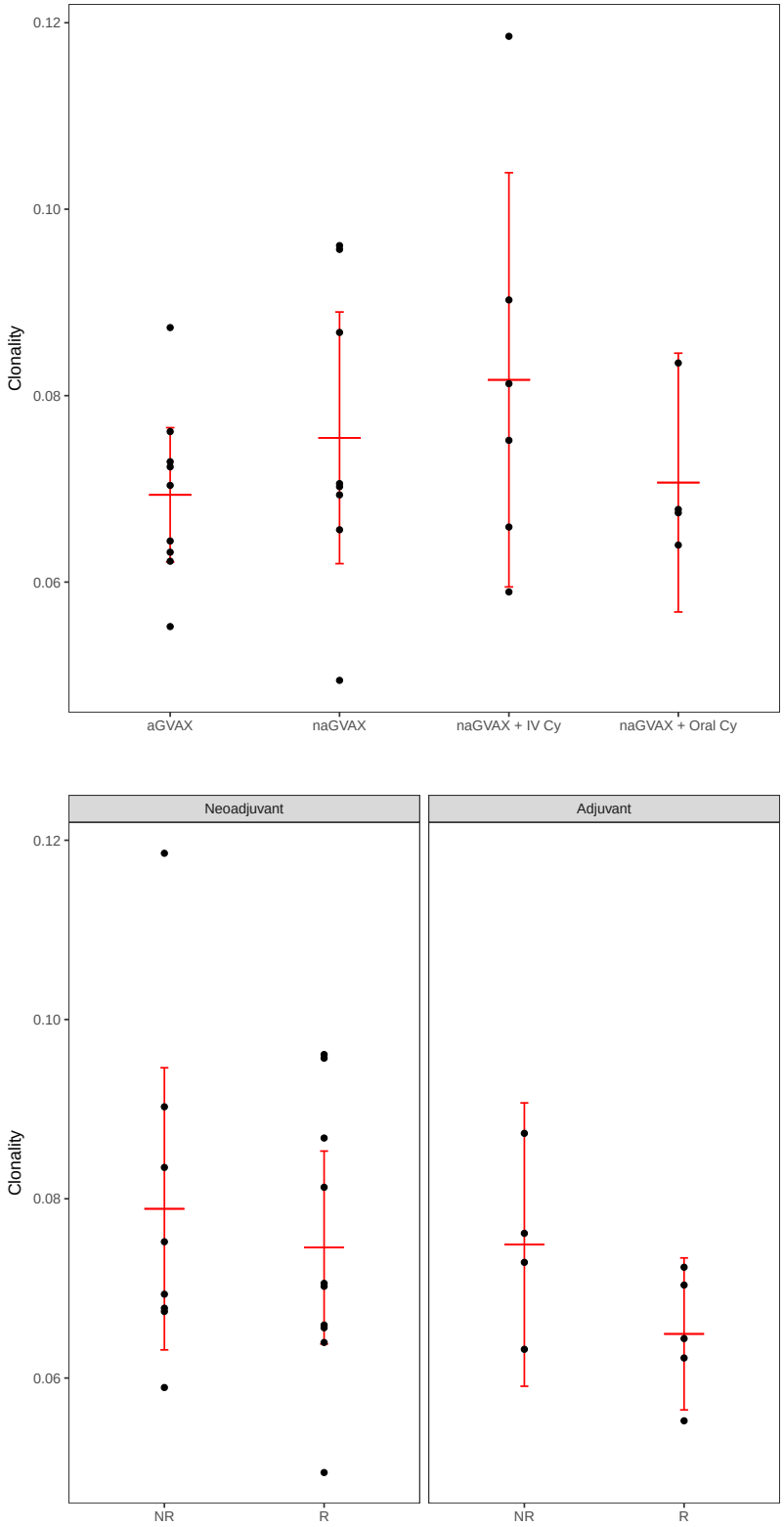

Figure 2.18: The clonality of tumor sections in the adjuvant and neoadjuvant trials, separated by treatment arm (top) and clinical response (bottom).

2.3.2 GVAX with Stereotactic Body Radiation Therapy

Patients enrolled in the J1179 clinical trial received Stereotactic Body Radiation Therapy (SBRT) after surgical resection, but before 6 courses of FOLFIRINOX chemotherapy (folinic acid, 5-fluorouracil, irinotecan and oxaliplatin). Patients in one arm also received cyclophosphamide and GVAX before radiation.

To assess any effects of GVAX treatment on SBRT treated patients, the clonality at each time point was compared between treatment arms. No significant differences in clonality were observed between GVAX treated and untreated patients at any of the time points (Figure 2.19).

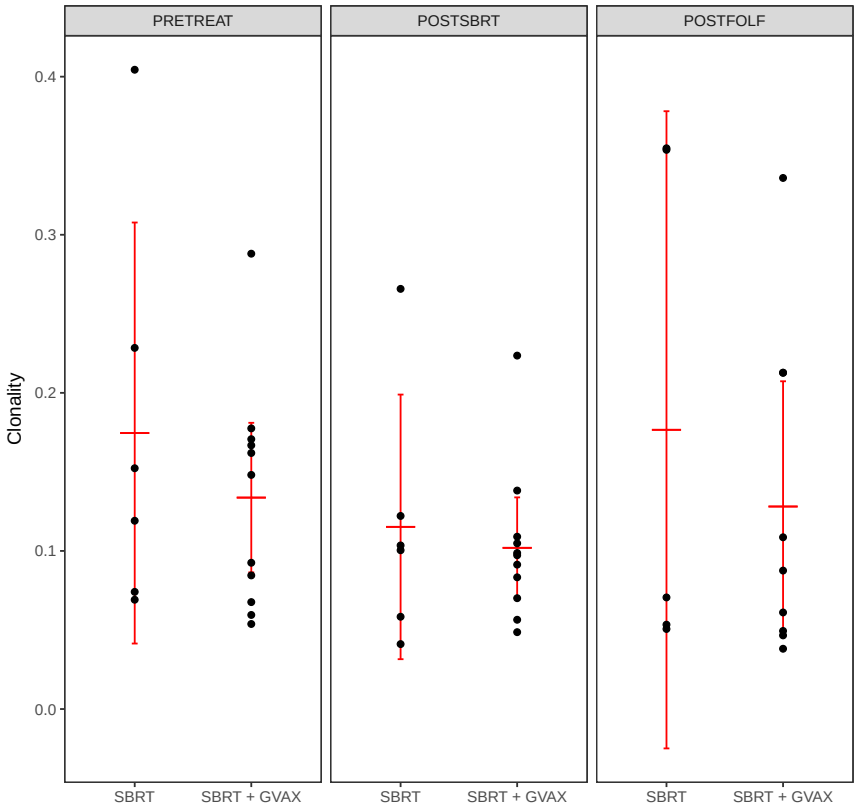

Figure 2.19: The clonality of all patients on the SBRT study before treatment (left) after radiation (center) and after chemotherapy (right), separated by treatment arm.

Comparing individual SBRT treated patients by fold change in clonality, no dif-

ferences were observed between treatment arms (Figure 2.20). Notably, all but three patients in this trial experienced a net diversification of their repertoires, unlike patients treated in the neoadjuvant study, in which most patients' repertoires became more clonal (Figure 2.6).

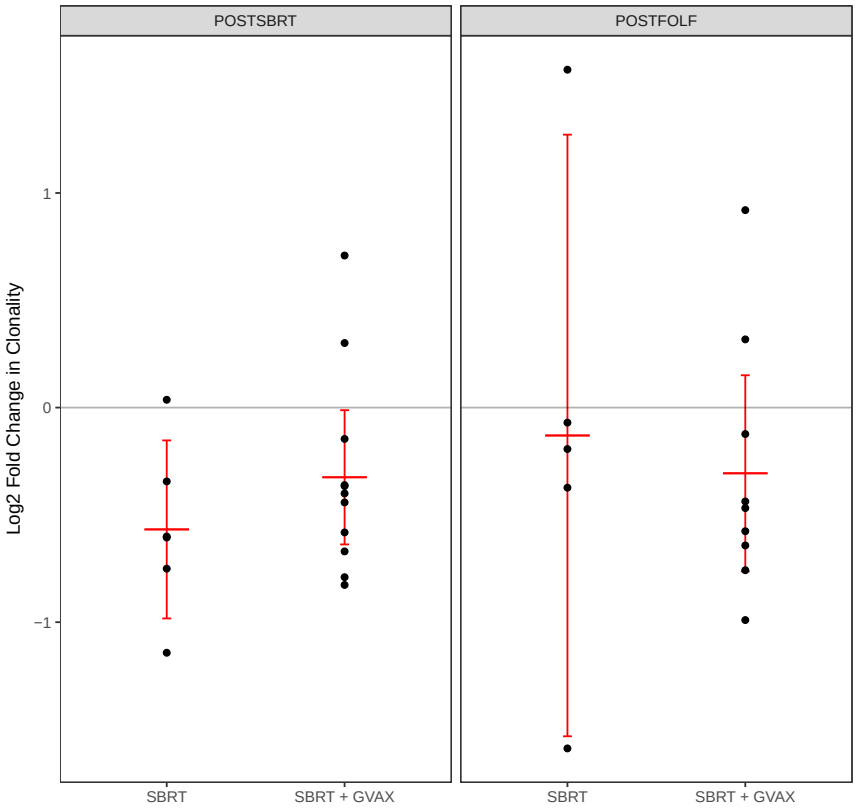

Figure 2.20: The Log2 normalized fold change relative to pre-treatment of all patients on the SBRT study, after radiation (left) and after chemotherapy (right), separated by treatment arm.

To assess if this decrease in clonality is consistent with T-cell death induced by radiation, a computational simulation of T-cell death was developed in which a set percentage of clones was randomly removed from the data and the clonality and richness were recalculated. The process was repeated 10000 times to generate a distribution of simulated post-radiation repertoire metrics. A representative baseline sample (with clonality 0.06, richness 158605) was selected, and 0.25%, 0.5% or 1% of

the clones were randomly chosen and deleted (this represented a loss of 675, 1350 and 2700 clones from the original 1.6×10^5, respectively). The distributions of the simulated post-radiation repertoires (Figure 2.21) demonstrated that deletion of clones from the repertoire always results in reduced richness, by necessity, but more interestingly it also always results in a decrease in clonality, with larger removals resulting in larger diversifications. This counterintuitive property arises from the fact that the majority of clones in a diverse sample are present at very small numbers, with a few clones existing in large quantities. Random deletions, therefor preferentially affect these clones with high abundance first. It is worth noting that the magnitude of the change in clonality demonstrated by these simulations is very small in comparison to changes seen in biological samples, indicating that massive loss of T-cells would be required to have a measurable effect on the clonality of a repertoire.

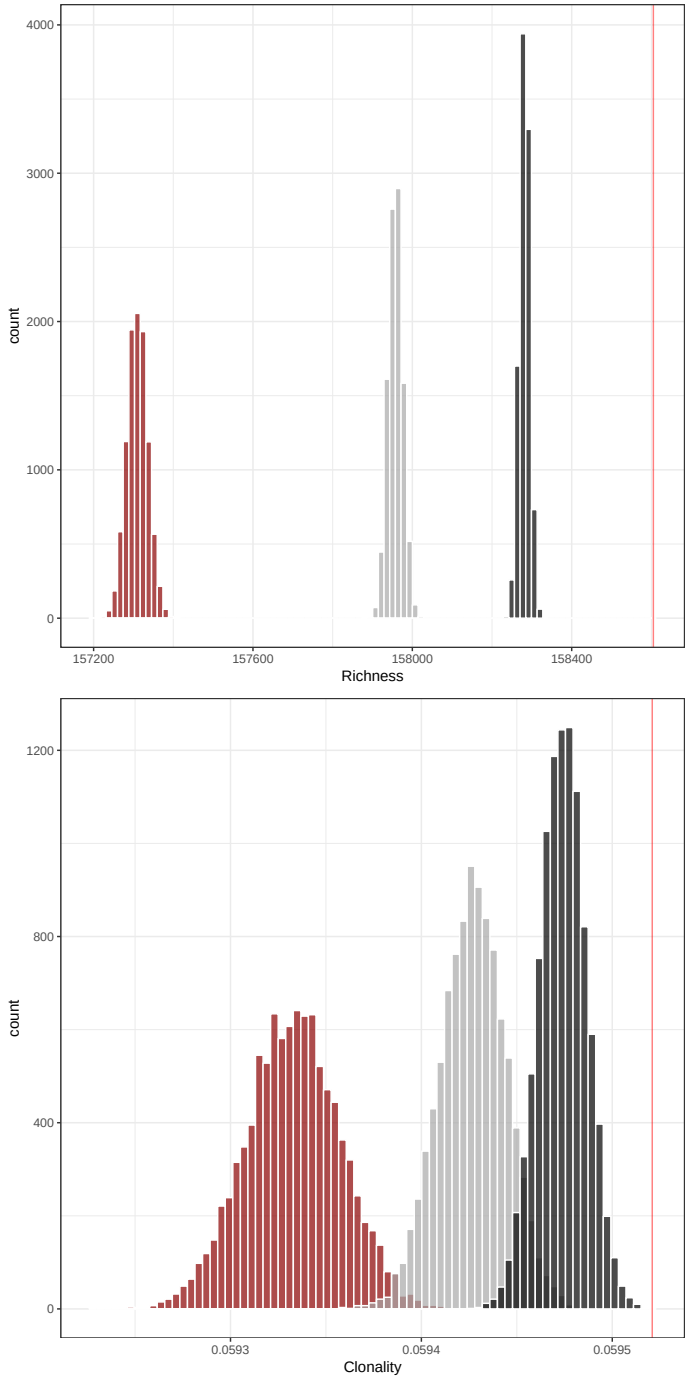

Figure 2.21: The distributions of richness (top) and clonality (bottom) in simulated post-radiation repertoires in which 1% (black), 0.5% (grey) or 0.25% (red) of clones were randomly deleted, 10000 times each. The red lines represent the richness or clonality of the pre-treatment sample from which the simulations were created.

Despite the changes observed in clonality, neither the number of expanded clones (Figure 2.22), nor the Morisita index (Figure 2.23) were significantly different between GVAX treated and untreated patients.

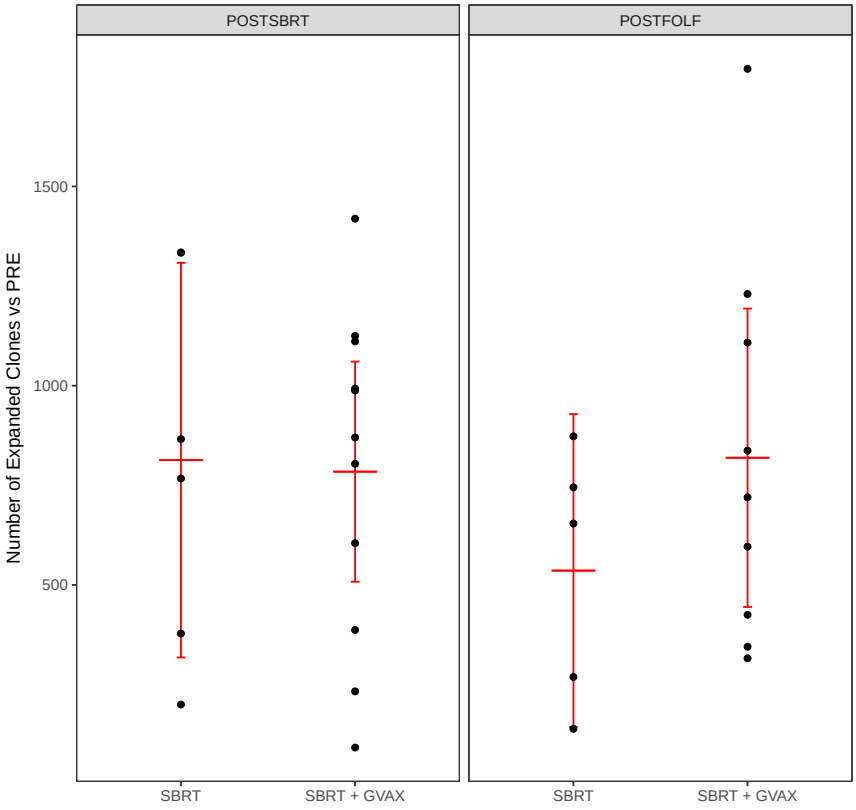

Figure 2.22: The number of significantly expanded clones relative to pre-treatment for all patients in the SBRT study, after radiation (left) and after chemotherapy (right), separated by treatment arm.

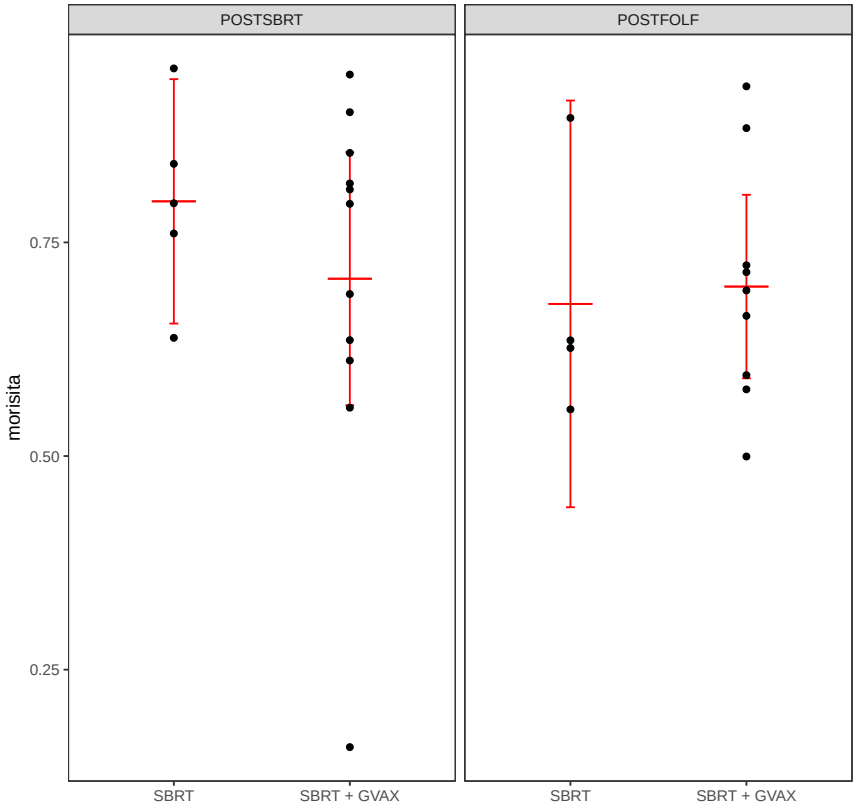

Figure 2.23: The Morisita index comparing all patients pre-treatment samples to time points after radiation (left) and after chemotherapy (right), separated by treatment arm.

To better understand the effects of targeted radiation on the peripheral immune system, T cell repertoire metrics were correlated with the reported Planned Treatment Volumes (PTV80) which measures the volume of tissue expected to receive significant amounts of radiation during treatment. No significant correlations were observed between PTV80 and clonality at any time point (Figure 2.24).

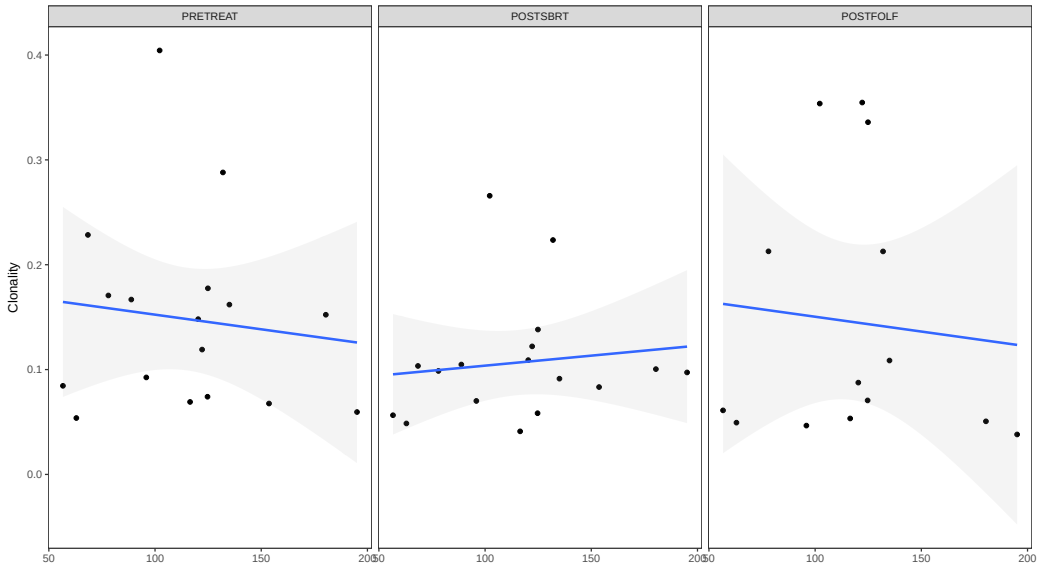

Figure 2.24: The clonality of all patients on the SBRT study before treatment (left) after radiation (center) and after chemotherapy (right), compared with the radiation volume (PTV80).

Using the log2 fold change of clonality relative to baseline, a significant correlation was observed after SBRT, with lower radiation volumes associated with greater diversification of the repertoire ($p = 0.0198$). Interestingly, this trend was completely reversed after treatment with FOLFIRINOX, although the negative correlation was not significant (Figure 2.25).

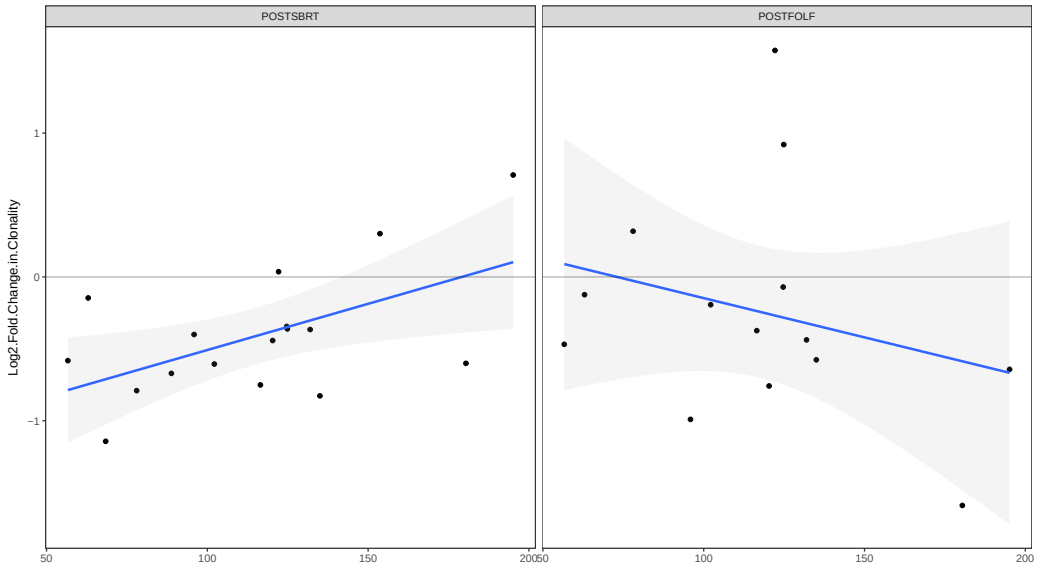

Figure 2.25: The Log2 normalized fold change in clonality, relative to the pre-treatment time point after radiation (left) and after chemotherapy (right), compared with the radiation volume (PTV80).

No significant correlations were observed between radiation volume and either the number of expanded clones (Figure 2.26) or the Morisita index (Figure 2.27).

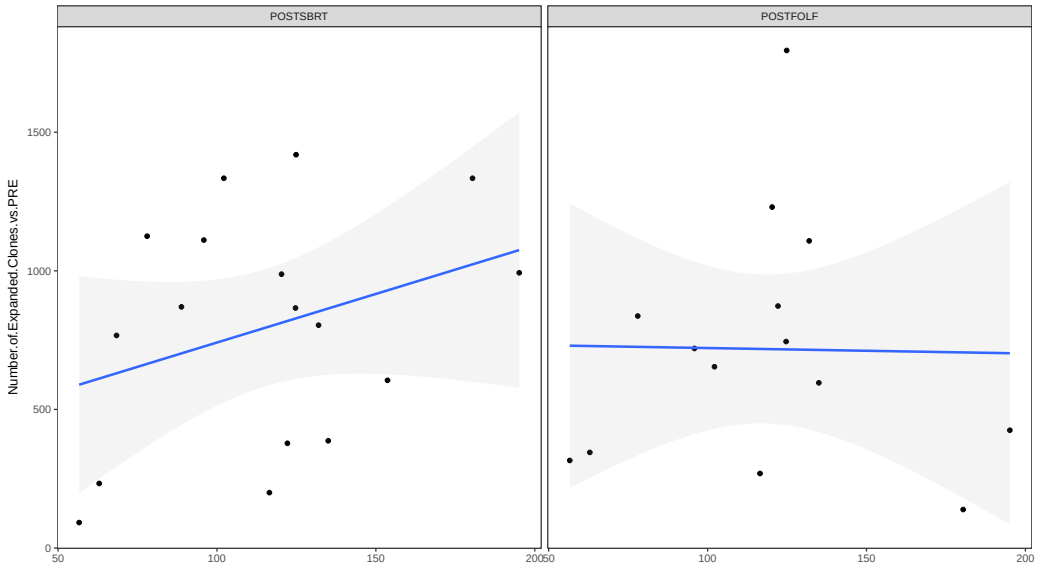

Figure 2.26: The number of significantly expanded clones after radiation (left) and chemotherapy (right) for patients in the SBRT study, compared with the radiation volume (PTV80).

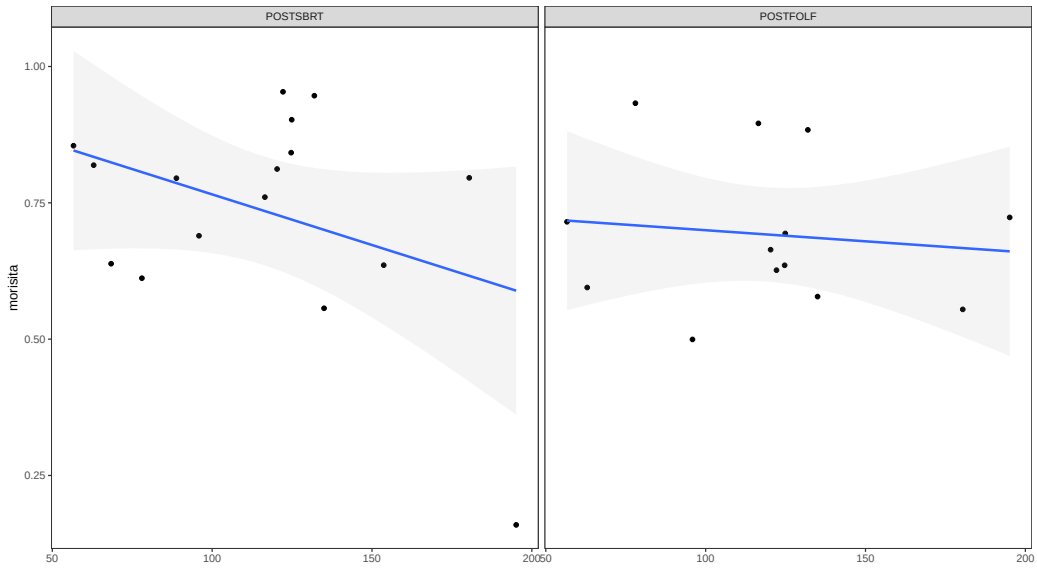

Figure 2.27: The Morisita index showing the change in the repertoire relative to pre-treatment, after radiation (left) and after chemotherapy(right), compared with the radiation volume (PTV80).

2.3.3 GVAX with Immune Checkpoint Inhibitors

Patients in the J0834 clinical trial received anti-CTLA4 with or without GVAX, while patients in the J14113 trial received GVAX followed by CRS-207 with or without anti-PD1.

To assess the baseline characteristics of patients in each treatment arm of these studies, as well as to examine the overall effect of treatment, the clonality was compared between treatment arms both before (Figure 2.28, left) and after (Figure 2.28, right) treatment. At baseline, the average repertoire clonality was slightly lower (more diverse) in patients enrolled on the anti-CTLA4 study, however this trend was not significant. For all treatment arms, peripheral repertoires were slightly more diverse after treatment, again not reaching significance for any arm.

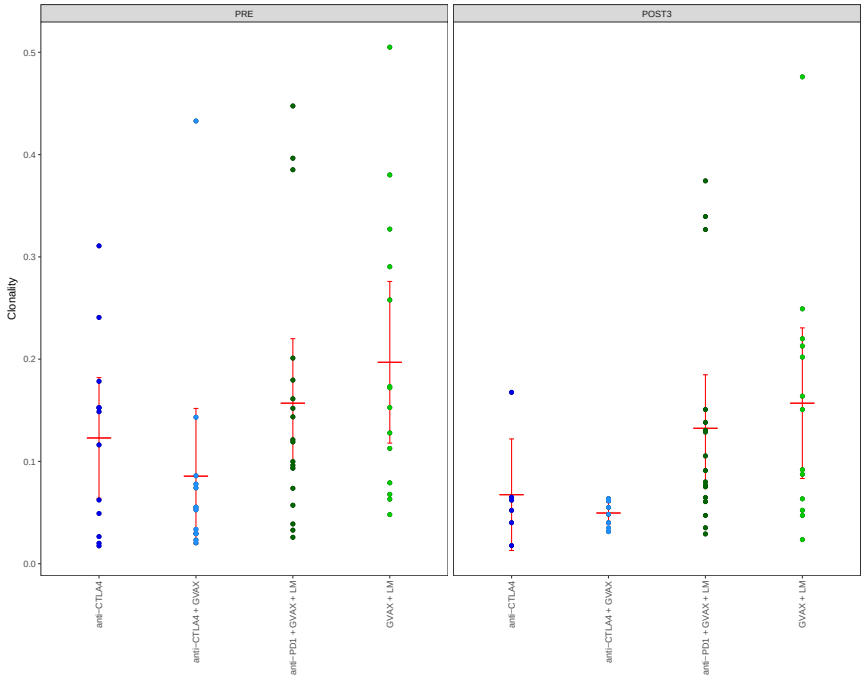

Figure 2.28: The clonality of all patients at baseline (left) and after 3 cycles of treatment (right), separated by treatment arm.

To examine the effect of treatment on individual patients, the log2 fold change in

clonality was compared between all treatment arms (Figure 2.29). The majority of patients experienced a net diversification of their peripheral repertoires. A greater proportion of the patients enrolled in the anti-PD1 trial experienced diversification.

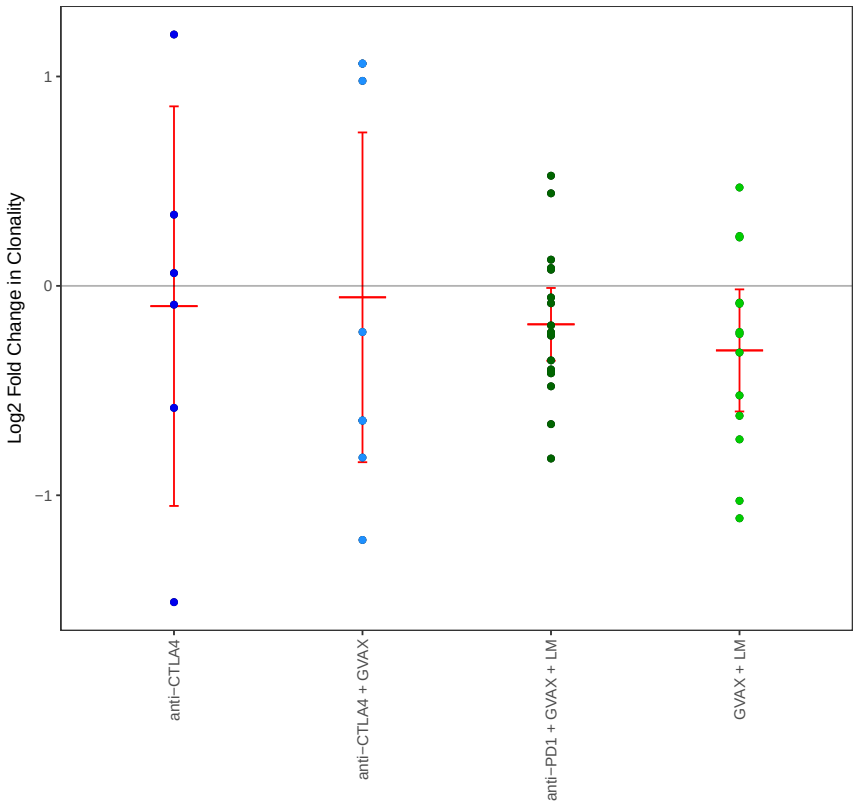

Figure 2.29: The Log2 normalized fold change in clonality relative to baseline for each patient on the anti-CTLA4 and anti-PD1 trials, separated by treatment arm.

No significant difference was observed in the number of expanded clones between treatment arms (Figure 2.30), however patients receiving anti-CTLA4 with GVAX had significantly larger changes in their repertoires than patients receiving anti-PD1 with GVAX and CRS-207 (Figure 2.31).

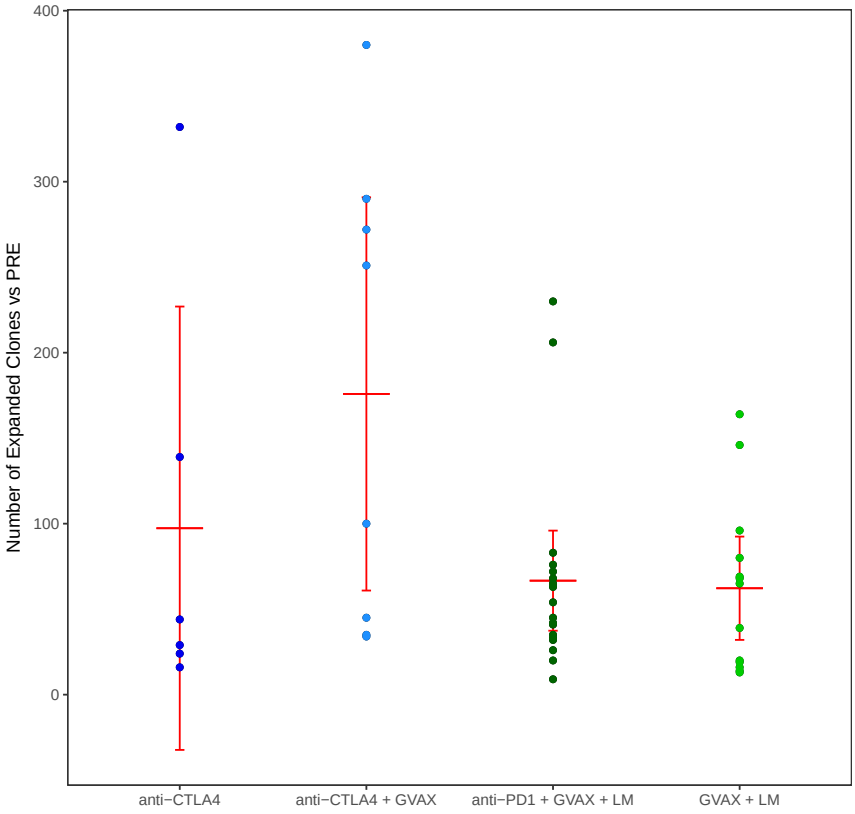

Figure 2.30: The number of expanded clones following treatment for all patients on the anti-CTLA4 and anti-PD1 trials, separated by treatment arm.

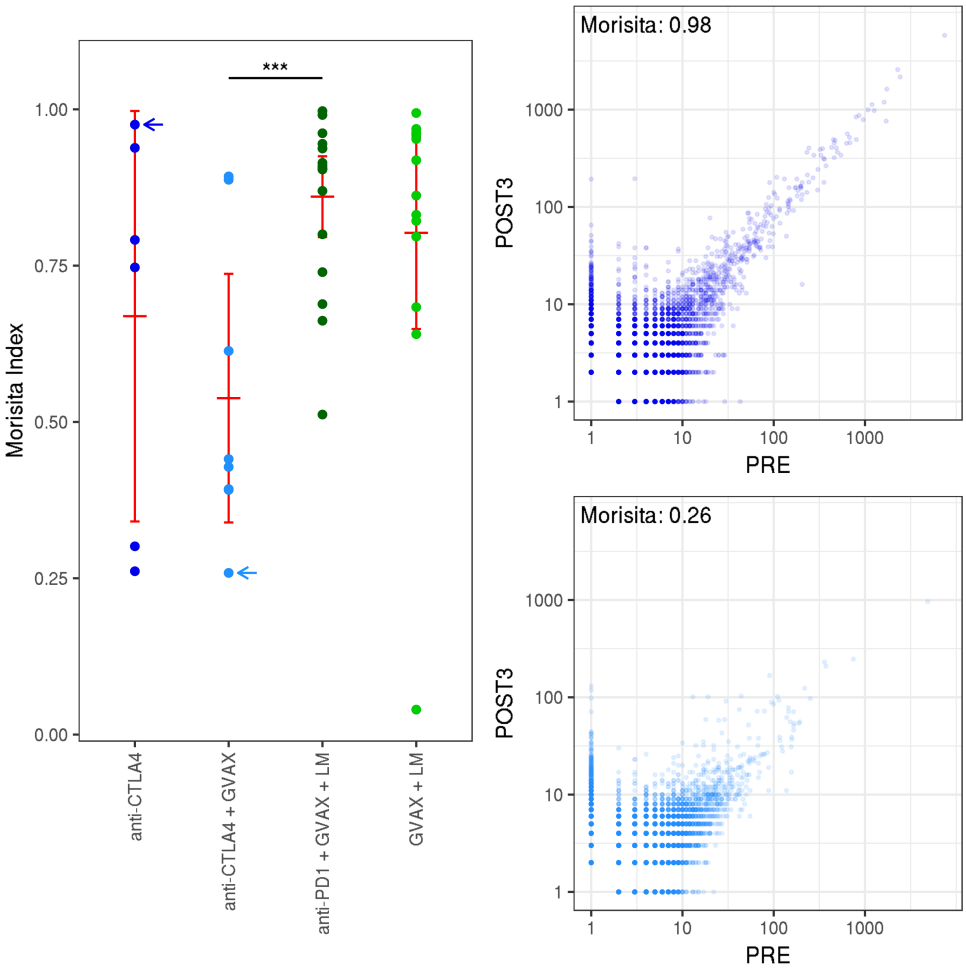

Figure 2.31: The Morisita index of all patients, separated by treatment arm (left). Representative scatter plots showing each clone before and after treatment for patients with high (top right) or low (bottom right) Morisita index. Arrows indicate which repertoires are shown at right.

Finally, to assess the ability of TCR repertoire metrics to differentiate clinical responders, each metric was used to compare patients with overall survival of > 6 months (R) or < 6 months (NR). At baseline, anti-CTLA4 treated patients with

longer overall survival had more diverse peripheral TCR repertoires (Figure 2.32). This trend was completely absent in patients in the anti-PD1 trial. Interestingly, the trend was no longer observed after treatment (Figure 2.32), possibly because most patients experienced a diversification following treatment (see Figure 2.29).

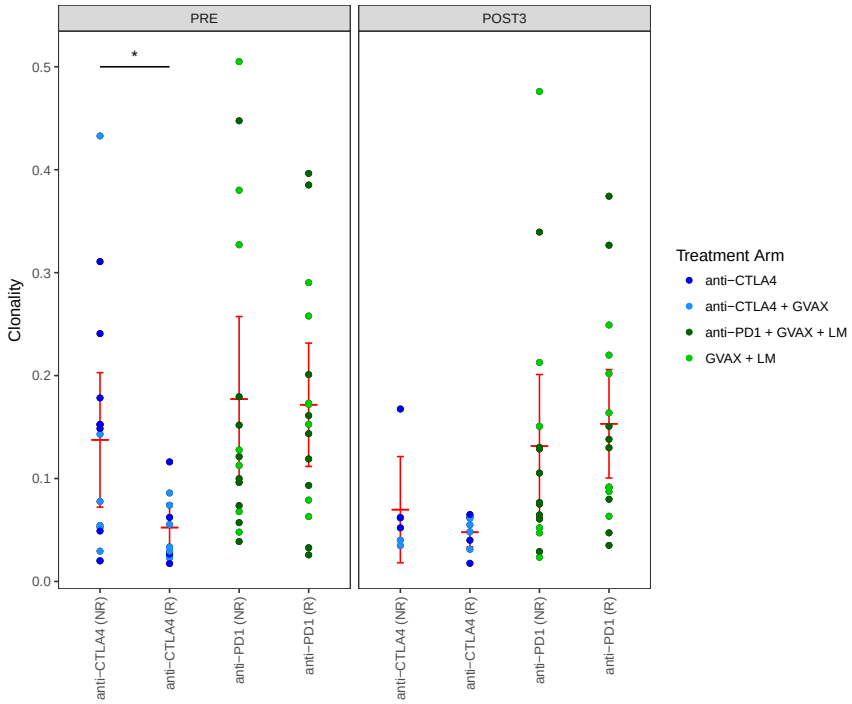

Figure 2.32: The clonality of each patient before (left) and after (right) treatment. Samples are separated by clinical response and trial, and colored by treatment arms.

Because of the striking difference in baseline repertoire clonality, the log2 fold change in clonality is more difficult to interpret. No significant differences were observed between responders and non-responders in either study using this metric (Figure 2.33). Non-responders seem to have slightly more diversification than clinical responders, however this is likely the result of extremely low baseline clonality in clinical responders, rather than a measurement of change.

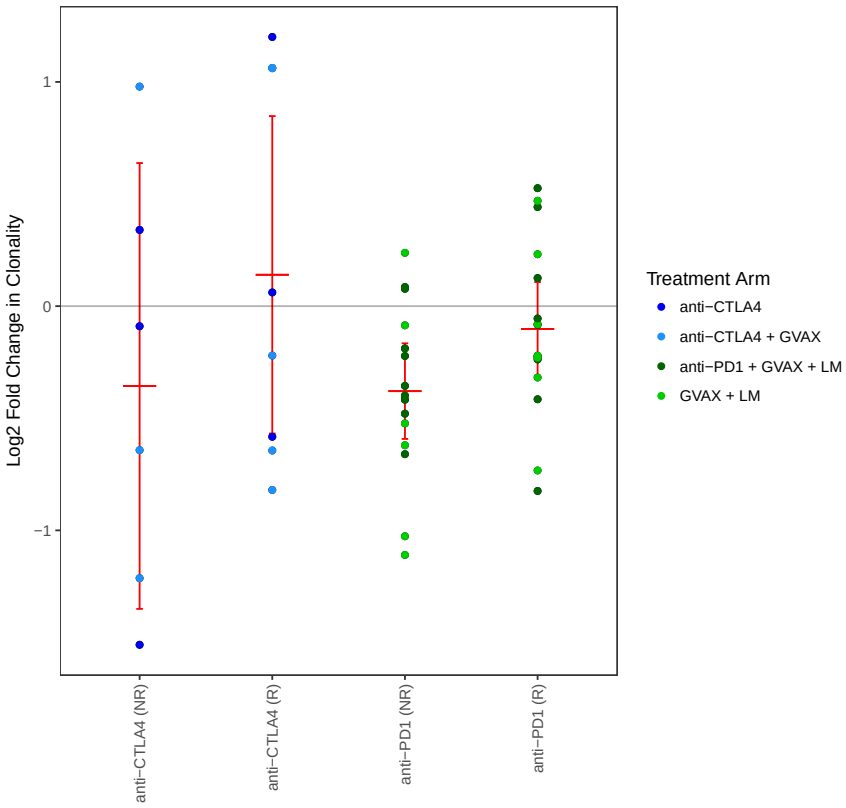

Figure 2.33: The log2 normalized fold change in clonality of all patients on the anti-CTLA4 and anti-PD1 trials, separated by trial and clinical response, and colored by treatment arm.

Clinical responders treated in the anti-CTLA4 study experienced dramatically more expanded clones following treatment (Figure 2.34), with mean 225.5 expanded clones in responders, compared with 33.8 in non-responders. As seen previously with clonality, the trend was not observed in patients receiving anti-PD1, where clinical responders averaged 60.2 expanded clones compared with 69.7 for non-responders.

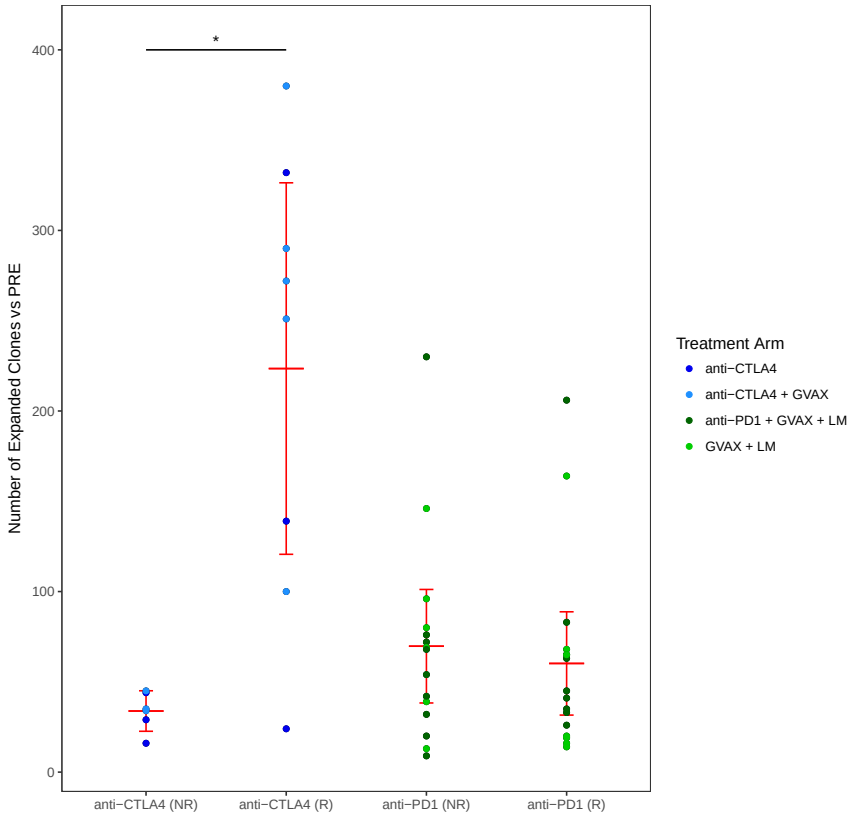

Figure 2.34: The number of expanded clones following treatment in patients enrolled in the anti-CTLA4 and anti-PD1 trials, separated by clinical response and trial, and colored by treatment arm.

The Morisita index was largely unchanged between clinical responders and non-responders in both studies, with patients in the anti-CTLA4 study showing a trend towards more change in patients with longer survival. None of the changes observed were significant using this metric (Figure 2.35).

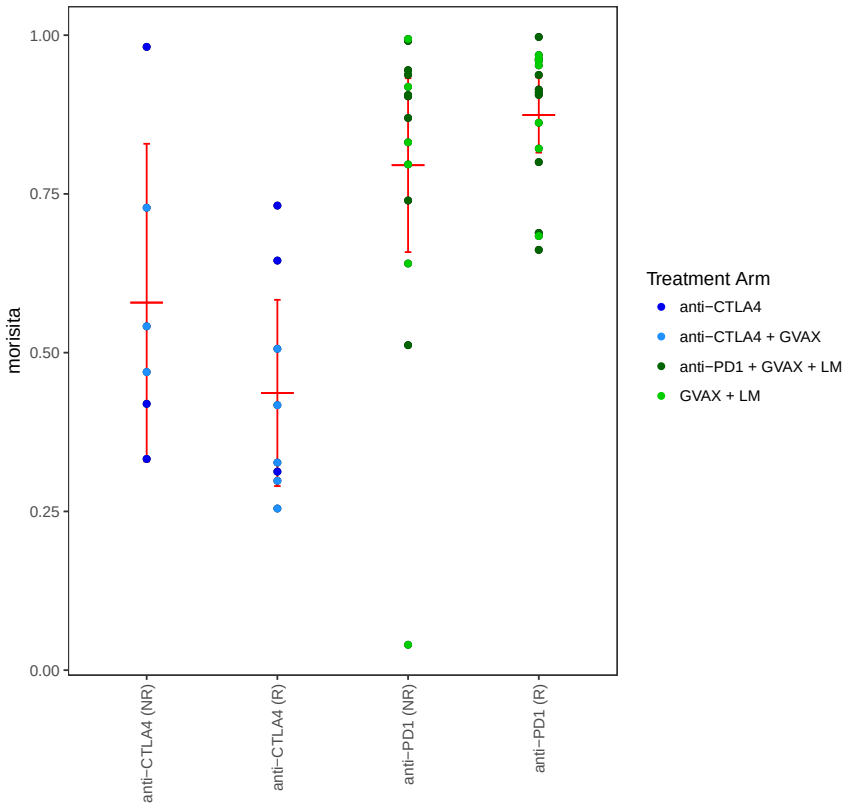

Figure 2.35: The Morisita index of all patients on the anti-CTLA4 and anti-PD1 trials, separated by clinical response and trial, and colored by treatment arm.

Because of the striking association between clinical outcome and both diverse baseline repertoires and high number of expanded clones, both were as potential biomarkers using this dataset. A Kaplan-Meier estimator was used to model survival in both studies, separating patients by either clonality (greater or less than 0.1) or number of expanded clones (greater or fewer than 100). Patients with diverse baseline repertoires survived significantly longer ($p < 0.05$) than patients with clonal repertoires when treated with anti-CTLA4. Likewise, for patients treated with anti-CTLA4, patients who expanded more than 100 clones following treatment survived significantly longer ($p < 0.01$, Figure 2.36, top). Understandably, neither of these predictors were valid for patients treated on the anti-PD1 study (Figure 2.36, bottom).

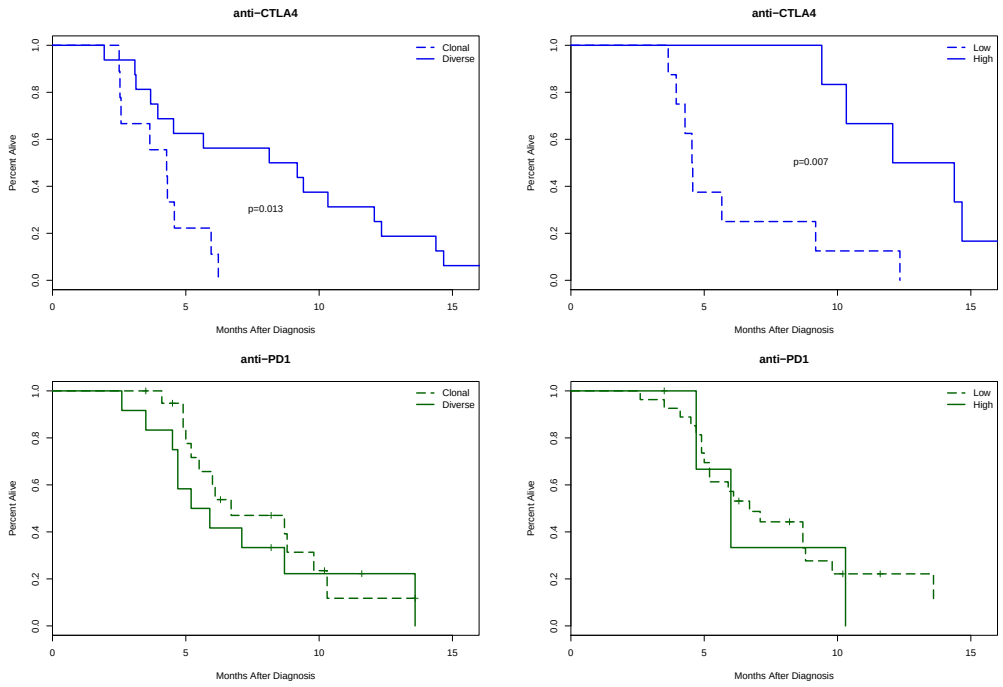

Figure 2.36: Kaplan-Meier survival analysis of patients enrolled in the anti-CTLA4 (top) or anti-PD1 (bottom) studies. Patients were divided into diverse or clonal groups (left, using 0.1 as a clonality cutoff) or high and low number of expanded clones (right, using 100 clones as a cutoff).

To compare the TCR metrics to other, more easily attainable possible biomarkers, the same analysis was performed using absolute lymphocyte count at baseline (greater or less than 1000), as well as the treatment arm. Patients with high baseline lymphocyte count survived marginally longer (not significant) than lymphopenic patients in both studies (Figure 2.37, left). Patient in combination therapy arms (anti-CTLA4 + GVAX and anti-PD1 + GVAX + LM) survived slightly longer than those in comparable arms (anti-CTLA4 only, or GVAX + LM only), but this trend was not significant (Figure 2.37, right).

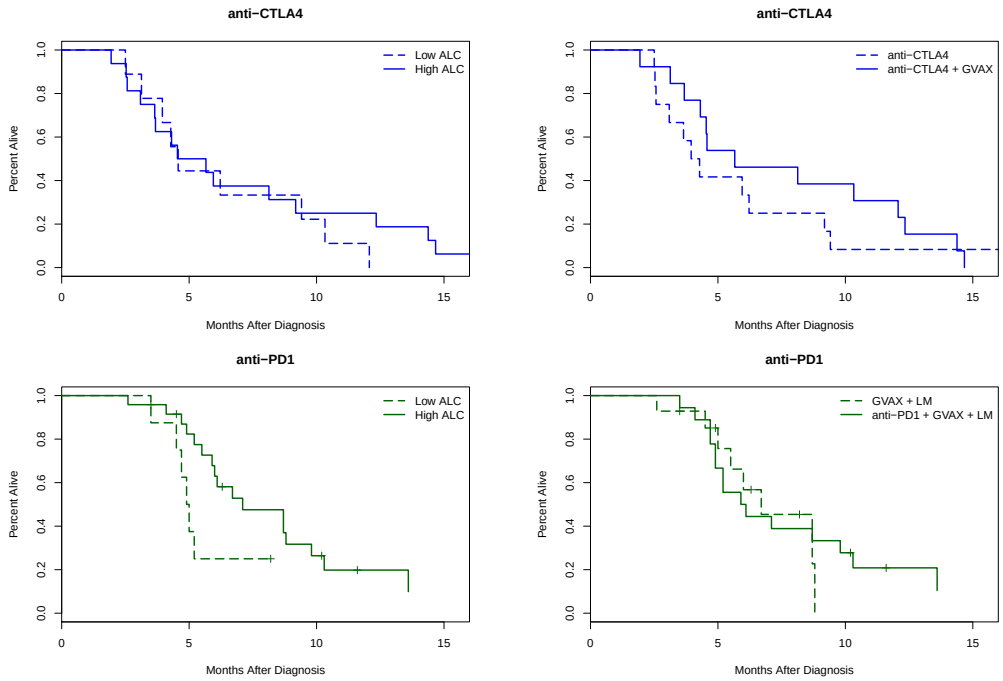

Figure 2.37: Kaplan-Meier survival analysis of patients enrolled in the anti-CTLA4 (top) or anti-PD1 (bottom) studies. Patients were divided into high or low absolute lymphocyte count (left, using 1000 as a cutoff) or treatment arm (right).

Because anti-PD1 acts at the site of the tumor, we next compared TCR repertoires of pre-treatment tumor biopsies to post-therapy biopsies in a subset of patients. The sample clonality is a valid representation of the total T-cell population unless the sample size is very small. To assure that the clonalities used in the tumor analysis were valid, any sample with fewer than 100 total TCR sequences was discarded. This resulted in two pre-treatment biopsies and three post-treatment biopsies identified as unusable for analysis (Figure 2.38). Unfortunately, the removal of these samples left only one sample in the GVAX + LM arm with paired pre and post-treatment samples, making group comparisons difficult.

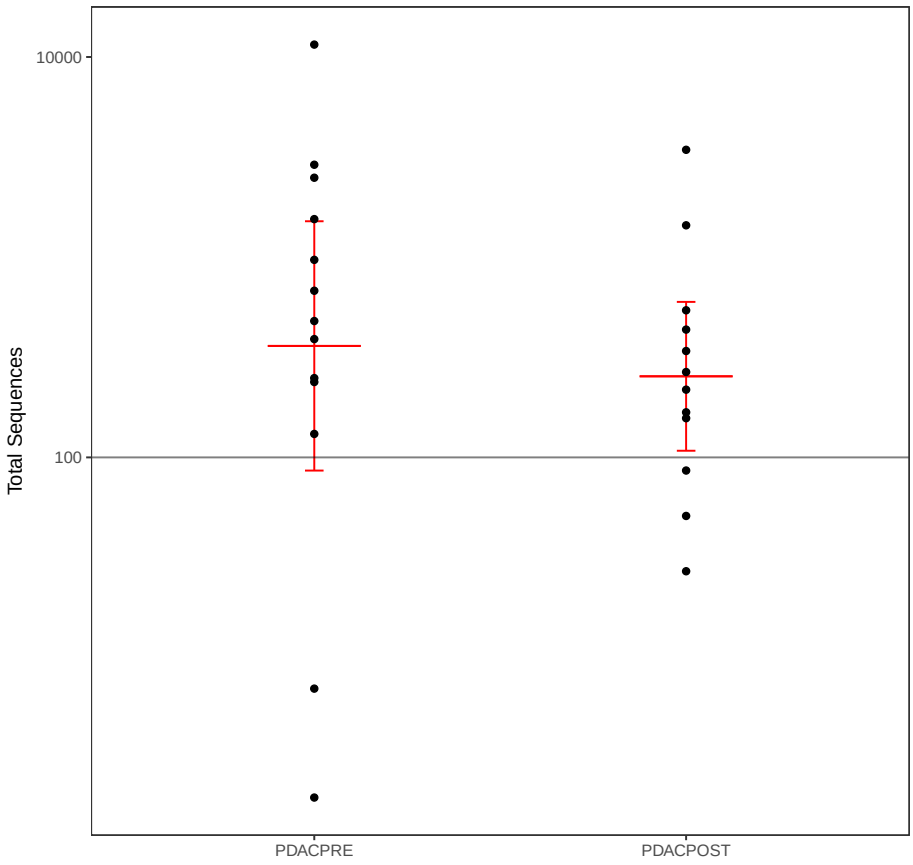

Figure 2.38: Total number of TCR sequences in tumor biopsies from the anti-PD1 trial

Patients who received anti-PD1 had no significant differences in their baseline repertoire clonality, and post-treatment clonality was incomparable due to low sample sizes (Figure 2.39). Metrics comparing pre and post-treatment samples (change in clonality, overlap, expanded clones) were not calculated due to low sample sizes.

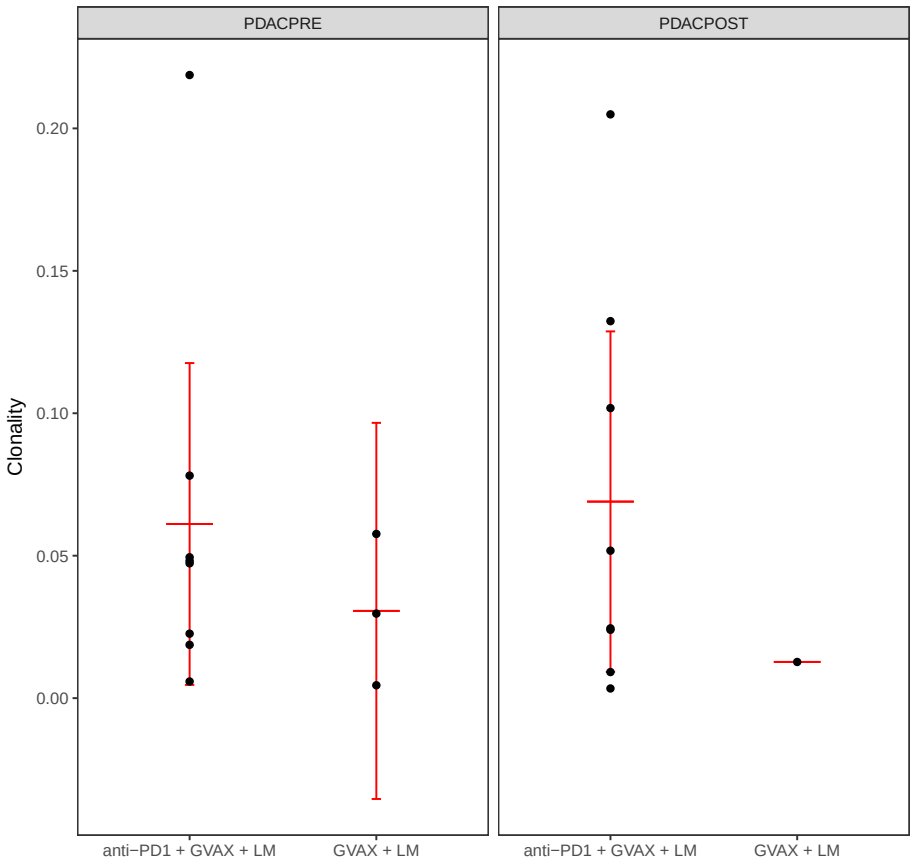

Figure 2.39: The clonality of tumor samples from patients in the anti-PD1 trial, separated by time point and treatment arm

After filtering, sufficient numbers of samples for comparisons between clinical responders and non-responders remained. No difference was observed between response groups in the clonality at either time point (Figure 2.40). Likewise, no trend was observed comparing these groups by the fold change in clonality following treatment (2.41).

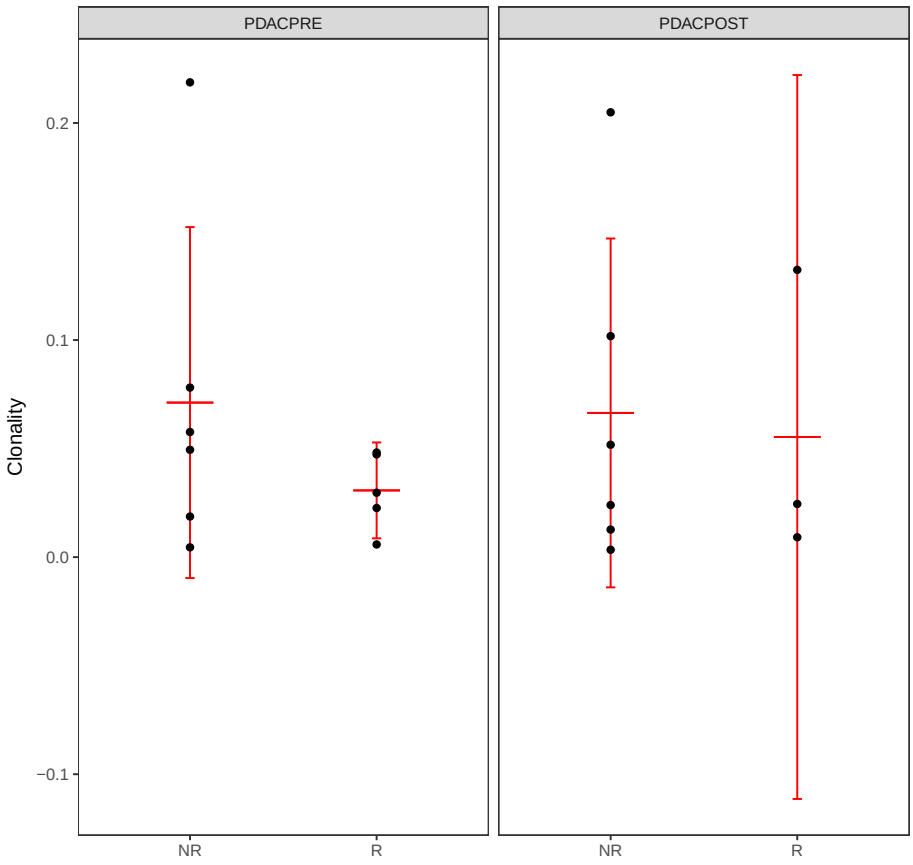

Figure 2.40: The clonality of tumor samples from patients in the anti-PD1 trial, separated by clinical response

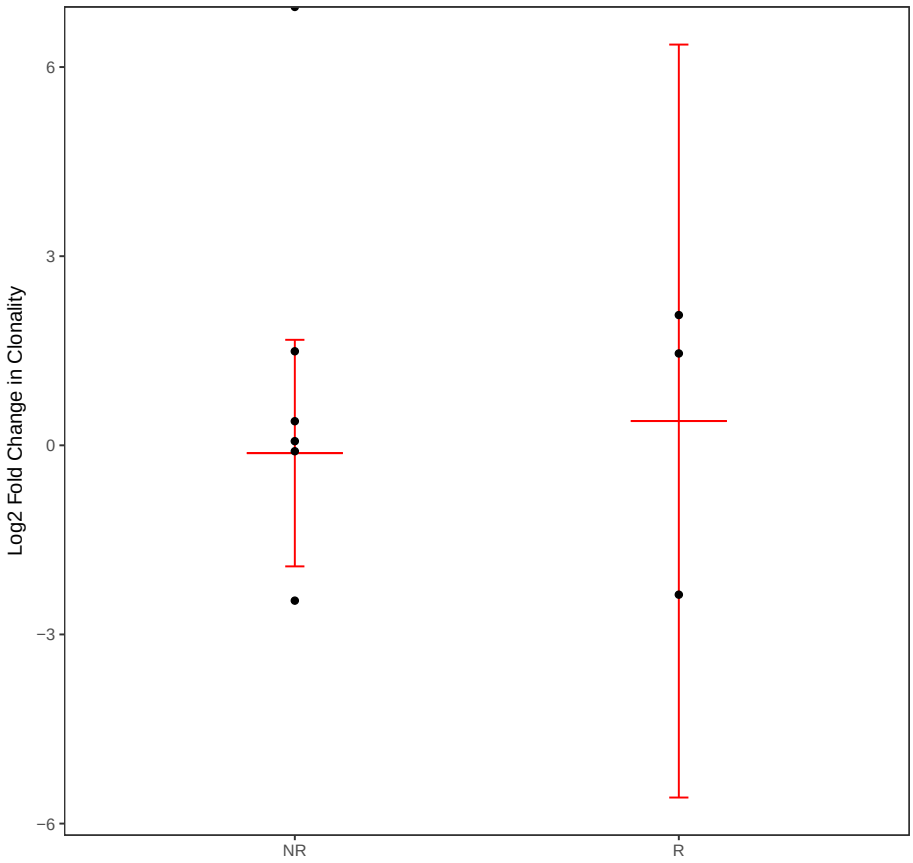

Figure 2.41: The Log2 fold change in clonality of the tumor repertoire, from patients in the anti-PD1 trial, separated by clinical response.

The Morisita index was unchanged between responders and non-responders (Figure 2.42), though the small sample size makes interpretation of these findings difficult.

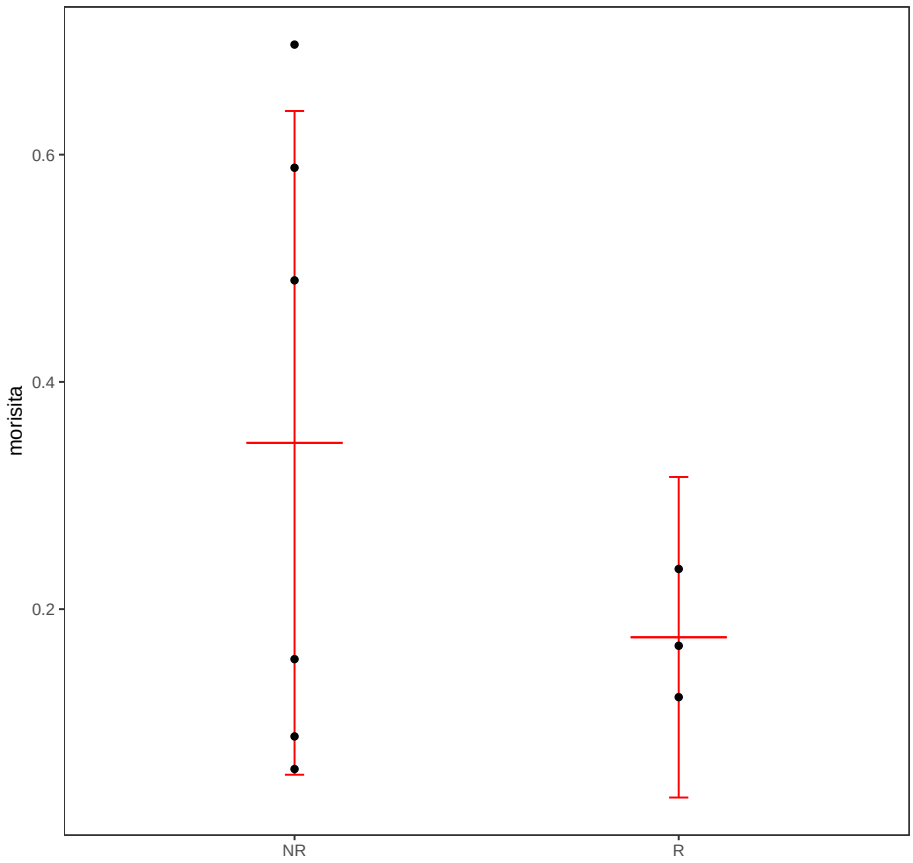

Figure 2.42: The Morisita index comparing pre and post-treatment samples from patients on the anti-PD1 trial, separated by clinical response

2.4 Discussion

This study represents the largest collection of TCR repertoire profiles from pancreatic cancer trials to date. Repertoire features were found to vary with treatment arm, and in some contexts, clinical response. These findings provide a platform for hypothesis generation as well as a starting point for the development of novel biomarkers in a disease which desperately needs them.

Analysis of the neoadjuvant and adjuvant GVAX trials demonstrated one of the main challenges associated with cross study analyses: patient heterogeneity. Because the baseline repertoire clonality was, on average, much lower in the neoadjuvant study, comparisons between the two studies were difficult to make. This difference in baseline characteristics is difficult to explain, as both studies recruited patients without metastatic disease and no significant difference was observed between the ages of the enrolled patients.

Patients treated with neoadjuvant GVAX experienced more changes in their repertoire following treatment than patients in the adjuvant trial. This could be a result of GVAX-primed T-cells interacting with the tumor and expanding following treatment. In the adjuvant study, the tumor is removed before vaccination, taking with it any existing population of tumor infiltrating T-cells, but also a large source of antigen for vaccine primed T-cells to attack.

Baseline differences also interfered with the ability to observe effects of cyclophosphamide on the T-cell repertoire. Because patients administered this drug had lower baseline clonality than those receiving only neoadjuvant GVAX, these arms appear to experience less diversification following treatment. This observation is likely an artifact which occurs in patients with very diverse baseline repertoires who have very little room for diversification following treatment.

The assay also failed to detect significant differences between clinical responders and non-responders, which may have been due to the small sample size. A power

analysis of the baseline samples from the adjuvant GVAX trial, for example, revealed that with the heterogeneity of the samples, the analysis only had 62% power to identify a difference in clonality of 0.085 (the difference separating clinical responders in the anti-CTLA4 study) with $n = 5$ samples per group. To achieve 80% power (with $\alpha = 0.05$), at least 8 samples per group would be necessary under the same assumptions.

TCR repertoire samples from the SBRT study showed no measurable difference between patients treated with GVAX and those only receiving SBRT and radiation. This study provided the first indication that the effect of GVAX on the peripheral repertoire is likely too small to be measured using these metrics. The study did support the hypothesis that radiation can damage the peripheral repertoire, as a clear correlation was observed between the change in clonality and the radiation volume, with the trend suggesting that lower volumes of affected tissue correspond to larger diversifications following treatment. It is worth noting that the magnitude of this change is relatively small, as no large shifts are seen in the average clonality before and after radiation treatment (Figure 2.19).

Finally, the immune checkpoint inhibitor trials demonstrated that the best application of TCR sequencing is in the context of treatments which have large effects on the peripheral repertoire. The anti-CTLA4 trial exhibited the largest bulk changes in the repertoire (Figure 2.31), and suggests (though it did not reach statistical significance) that the addition of GVAX may further enhance this change. GVAX is likely providing a large source of antigen which can be presented to T-cells by APCs (signal 1), while anti-CTLA4 allows efficient costimulation (signal 2). This mechanism also accounts for the extremely high number of expanded clones identified in a subset of patients enrolled on this trial.

Both diverse baseline repertoires and high number of expanded clones were significantly associated with clinical response in the anti-CTLA4 trial. Other groups have

demonstrated the potential of repertoire diversity as a biomarker for immunotherapy efficacy[20], but this is the first demonstration in pancreatic cancer.

The failure of these biomarkers in the context of anti-PD1 treatment is unsurprising, given that this treatment acts on tumor infiltrating lymphocytes. Other groups have noted the differences in peripheral repertoire metrics between CTLA4 and PD1 blockade therapy[22]. Data from tumor samples was inconclusive due to the small sample size after quality control filters were applied. The variance of clonality in the tumor samples was actually lower than that of blood samples, indicating that if the sample sizes had been larger, smaller changes in tumor clonality would be detectable. Unfortunately, surgical resections and biopsies were often unattainable and all available samples were used for this analysis.

The large numbers of samples used in the analysis of these four clinical trials provide a robust measure of the distribution of repertoire clonality in patients with pancreatic cancer (shown in Figure 2.43, left). Accurate description of this distribution allows a power analysis to be performed, approximating the number of samples needed to detect a given change in clonality under various assumptions (Figure 2.43, right).

This analysis of 343 TCR repertoire samples from 106 pancreatic cancer patients enrolled in four immunotherapy clinical trials has provided a number of new insights into the treatment of this disease, and the application of this technology. First, it can be concluded that the effects of cancer vaccines are difficult to detect at the level of the repertoire, while effects of immune remodelling therapies like anti-CTLA4 are more obvious. Second, it is clear that the clonality and number of expanded clones are important metrics in the analysis of TCR repertoire data from clinical samples. Finally, it is clear that in certain clinical contexts, these metrics have the potential to serve as biomarkers for clinical response.

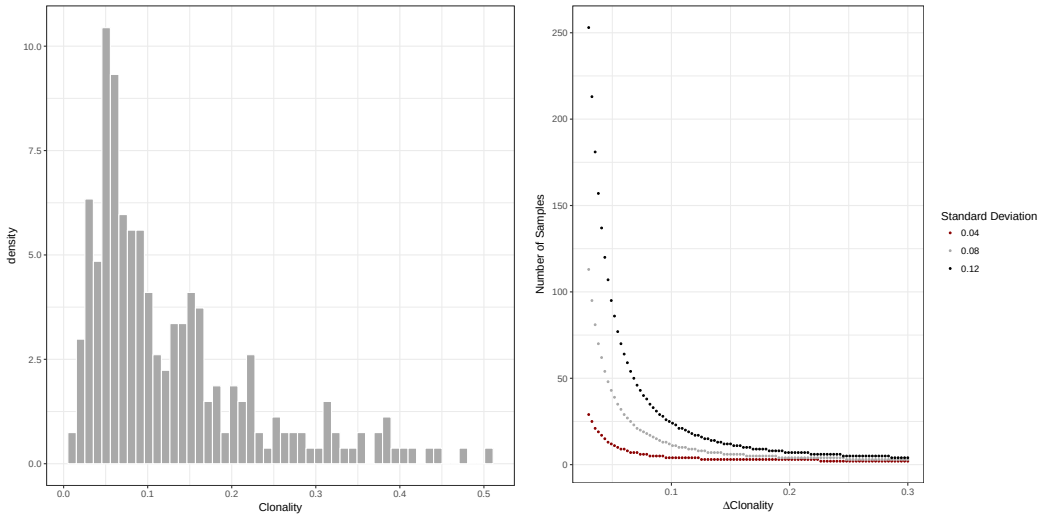

Figure 2.43: The distribution of the clonality from all peripheral repertoire samples used in this study ($n = 265$, left). A parametric power analysis using three assumptions for the standard deviation, showing the number of samples needed (vertical axis) to detect a given change in clonality (horizontal axis) with 80% power at $\alpha = 0.05$.

Chapter 3

Fusion Based Preventative Vaccines

3.1 Introduction

The next generation of cancer vaccines will be tailored to the unique set of mutations present in the tumors of each patient (the tumor mutome). Unlike GVAX and other allogeneic vaccines, personalized vaccines will target only mutations known to be present in the tumor, targets which will ideally be the result of a rational design process to select tumor antigens most likely to lead to immunity. A variety of methods have been developed to formulate vaccines targeting specific antigens, including genetic engineering of dendritic cells[30] or *Listeria monocytogenes*[12] to express tumor antigens. The simplest approach, however, is to administer peptides encoding tumor antigens directly. The peptides must be co-administered with an adjuvant to enhance antigen uptake and presentation by APCs, and induce costimulatory molecules and cytokines[31]. Peptide vaccines have been used successfully in preclinical models[32,33], but more recently they have been studied in human trials in head and neck[34], breast[35], colorectal[36], brain[37] and even pancreatic cancer[38] with varying degrees of success.

These majority of peptide vaccine trials target point mutations, often commonly occurring in their respective cancers, but other groups[39] have recognized the potential for targeting fusion peptides, which result from genomic rearrangements common in many cancers. Fusion-derived antigens offer a number of advantages as potential targets for peptide vaccines. Relative to point mutations, fusions generate more exotic peptide sequences which may be more likely to be recognized as foreign by the immune system. Chromosomal rearrangements frequently occur in introns (which are, on average, much larger than exons), which results in fusion products generated by exons from different genes which are spliced together. This results in patients with a wide array of different genomic rearrangements expressing the same fusion protein.

The EML4-ALK fusion is present in a small subset of non-small cell lung cancer. Around 15 variants of the fusion protein have been identified, but the three most common make up 71% of the cases[40]. The p53 tumor suppressor gene, in contrast, is mutated in around half of all cancers, but so many different mutations are possible that the three most common mutations are present in only 18% of the total cases[41]. Because of this disparity, vaccines derived from fusion antigens may be effective in a greater percentage of the target population.

Peptide vaccination efforts have benefited greatly from the development of algorithms capable of predicting peptide/MHC binding affinities. While not the entire picture, peptide loading is one of the biological bottlenecks of antigen presentation, and the ability to filter out vaccine candidates with little likelihood of presentation has dramatically improved design efforts. Early efforts to predict this complex molecular interaction were codified in the SYFPEITHI algorithm[42], which built a large database of known peptide/MHC affinities. More recently, an artificial neural network, NetMHC has been trained on the initial data from the SYFPEITHI database, augmented with other data sources[43]. The same group is developing tools to predict immunoproteosomal processing[44] and TCR recognition[45].

In this study, fusion proteins were identified and targeted in two preclinical murine models of lung and colon cancer. Fusion identification was achieved using RNAseq data, but validation efforts using whole genome sequencing showed that current algorithms to identify these events from next-generation RNA sequencing data result in unacceptably high false positive rates. The methods used to test immunogenicity of candidate peptides were successful, and the peptide vaccination strategy succeeded in a tumor prevention model. These successes demonstrate that with improved accuracy of rearrangement finding methods, fusion-based peptide vaccines may provide clinical benefit to patients with a variety of cancer types.

3.2 Materials and Methods

3.2.1 Cell Lines and Mice

All mice used in this study were C57/BL6, aged 5-6 weeks, and were purchased from Jackson Laboratories. Mice were treated in accordance with Institutional Animal Care and Use Committee and American Association of Laboratory Animal Committee approved policies.

We chose to use two well studied cell lines derived from the BL6 mouse, capable of growth when injected subcutaneously. MC38 is a chemically induced colon adenocarcinoma cell line[46], and Lewis Lung Carcinoma (LLC) is a lung carcinoma cell line derived from a spontaneous tumor in a BL6 mouse[47].

Both cell lines were maintained in DMEM with 10% Fetal Bovine Serum, 1% Penicillin/Streptomycin and 1% L-Glutamine. Cells were passaged in a 10% CO_2 incubator, and were washed with sterile PBS before injection.

3.2.2 Identification of Fusion Proteins

RNA was extracted from 5×10^6 cells of each cell line using the Qiagen RNeasy Kit. Samples were sequenced using an Illumina HiSeq using the default paired-end protocol for RNA-Seq. Reads were aligned to the mouse mm9 genome (accessed from the UCSC Genome Browser[48]) using the `bowtie` aligner[49]. Fusions were identified using the `tophat-fusion` algorithm[50] with the settings listed in Table 3.1.

Option	Name	Value
Inner mate distance standard deviation	`mate-std-dev`	80
Maximum intron length	`max-intron-length`	100000
Minimum fusion separation	`min-fusion-dist`	100000
Minimum base pair mapping	`fusion-anchor-length`	13

Table 3.1: Settings used for the `tophat-fusion` algorithm

Fusions identified with `tophat-fusion` were then manually curated by examining the read mappings in the Integrated Genomics Viewer[51] (IGV), Fusion events which could not lead to a protein product (i.e. loss of start codons) were discarded.

A 20 base pair sequence with the fusion at the center was extracted for each candidate fusion. Binding of candidate fusion peptides to T2-D^b and T2-K^b was predicted computationally using the `NetMHC 3.4` algorithm[52] by submitting all possible 8, 9, 10 and 11mer peptides contained in the total 20mer.

3.2.3 Validation of Candidate Immunogenicity

Candidate 20-mer sequences were synthesized by Peptide 2.0, with a target purity of 95% and stored frozen in lyophilized form until use.

To test the immunogenicity of fusion peptides, 5-6 week old BL6 mice were vaccinated with either 50μg of each peptide in PBS with 10μg of Poly I:C as adjuvant, or a control vaccine with the Ova peptide (GLEQLESINFEKLTEWTSS) and adjuvant. A boost vaccine of the same formulation was administered 7 days after the initial vaccine. On day 14, spleens were harvested and pooled by treatment group. Cell suspensions were created by forcing spleens through a 40μm screen into CTL media (RPMI, 10% FBS, 0.5% L-glutamine, 1% Penicillin/Streptomycin, and 0.05mM 2-mercaptoethanol). Red blood cells were removed by lysis using Ammonium-Chloride-Potassium (ACK) buffer. T cells were harvested by negative selection from whole splenocytes using the Dynal $CD8^+$ Negative Isolation kit.

Immunogenicity of the vaccine peptides was determined using an ELISpot assay. T2-D^b or T2-K^b cells were pulsed with 2μg/mL of each candidate peptide for 2 hours. $CD8^+$ T cells from treated mice were combined in equal parts with either pulsed T2-D^b or T2-K^b (10^5 of each) in triplicate wells of an anti-mouse-IFN-γ coated ELISpot plate (mAb AN18, Mabtech). To address the possibility of $CD4^+$ responses, as well as any response not restricted to D^b or K^b, 10^6 splenocytes were plated in triplicate with

2μg/mL of each peptide. After 12-18 hours the cells were dumped off and the plate was washed 6 times in PBS with 0.05% Tween. A biotinylated anti-IFN-γ secondary antibody (mAb R4-6A2, Mabtech) was added at 10μg/mL for 2 hours, and the plate was washed as before. An avidin-conjugated peroxidase (Vectastain ELITE ABS, Vector Laboratories) was added and incubated for 1 hour. The plate was washed and AEC substrate was added. The substrate was allowed to incubate for 15 minutes before it was removed by rinsing in tap water. Plates were allowed to dry and IFN-γ spots were counted using an ImmunoSpot analyzer.

3.2.4 Assessment of Anti-tumor Activity

Mice were vaccinated as described in Section 3.2.3, and on day 14 after initial treatment, they were challenged with 10^5 tumor cells, injected subcutaneously into the leg. Tumors were measured using calipers in two perpendicular directions and the resulting measurements were averaged and reported as the average tumor diameter.

In tumor measurement studies, an additional vaccine group was treated using 3 peptides previously described to induce an anti-tumor response in the MC38 model[53]. These peptides were derived from the *Reps1* gene (VLELFRAAQLANDVVLQIME), the *Adpgk* gene (VHLELASMTNMELMSSIVHQ) and the *Dpagt1* gene (LVISASIIVFNLLELEGDYR).

3.3 Results

3.3.1 Identification of Fusion Proteins from RNAseq

In the MC38 cell line, 20,593,925 paired tags were aligned to the mouse `mm9` genome by `bowtie`. `tophat-fusion` Identified 14 putative fusions in this dataset, and inspection using `IGV` identified 8 which could potentially yield a fusion protein product.

The Lewis Lung Carcinoma sample was similar, with 21,419,009 paired tags aligned to the genome. Nineteen fusions were identified, with 9 potential, translatable fusions found.

Name	Amino Acid Sequence	NetMHC
LLC-1	ESHLRTSVLLPLRQTHASEP	
LLC-5	VPYRLELDHKVTMALLGKHC	
LLC-6	IRTCTDIKPEWLVKIAPPPN	
LLC-10	RTAPDYFLLQVLLKFHTDKG	H-2-Kb
LLC-11	TSPATPLPAGSGP	
LLC-16	LLEVLSGDSLARGHDPSAHV	
LLC-17	VPVPDISQEPERS	
LLC-18	GSPSFLKQKYGAGYHMTLVK	H-2-Kb
LLC-19	ADEATTPQAKKGRKKGIFPF	
MC38-1	EPRIEIVVVGTGT	
MC38-4	RLTLADKENTPPTLSSTGIL	
MC38-5	IRTCTDIKPEWLVKIAPPPN	
MC38-7	FRQSLSCCVCGGLPKTRYIQ	
MC38-8	DHGMEGNNTAEHS	
MC38-10	RGVSTVTDAASGTFDRSVTL	H-2-Kb
MC38-12	QYTGLDDHLSKYQNP	
MC38-13	AVMGRTLSMTLKHLESSSLL	H-2-Kb

Table 3.2: Candidate fusions identified by RNAseq, along with an indication of which peptides had <500nM predicted affinity for MHC alleles.

3.3.2 Validation of Candidate Immunogenicity

Using `NetMHC`, 4 peptides were identified as having at least moderate binding affinity (<500nM) to H-2-K^b. The sequences, names used in this study, and NetMHC binders are identified in Table 3.2.

To determine which peptides were capable of inducing an immune response, an ELISpot assay was performed using CD8$^+$ T cells from vaccinated mice as effector cells and T2 cells expressing mouse MHC molecules (T2-D^b or T2-K^b) as antigen presenting cells. Four peptides derived from LLC induced strong IFN-γ secretion (LLC-6,10,16 and 19) when presented by T2-K^b . Two of these peptides (LLC-16 and 19) also induced some response when presented by T2-D^b , which may be the result of presentation by the human HLA-A2 molecule, which is also present on the T2 cell line (Figure 3.1).

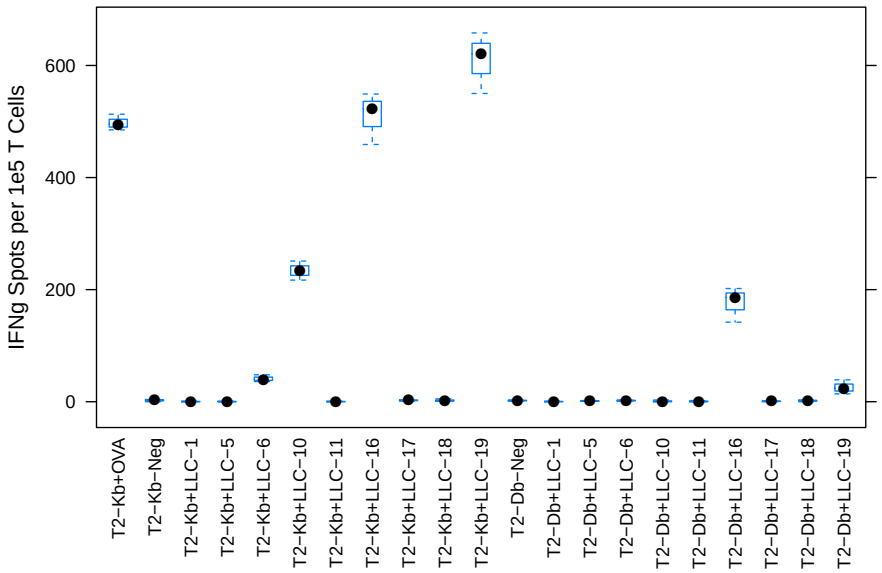

Figure 3.1: IFN-γ ELISpot results from the LLC candidate fusions, using T2-K^b and T2-D^b as antigen presenting cells. Ova, used as the positive control included in the vaccine.

To address the possibility of antigen presentation by HLA-A2, the assay was

repeated using T2 cells expressing neither mouse MHC molecule (but still expressing HLA-A2). The T2 cell line was able to present LLC-16 and 19 using HLA-A2, however dramatically fewer spots were seen, indicating that the peptides may be presented on both molecules (Figure 3.2).

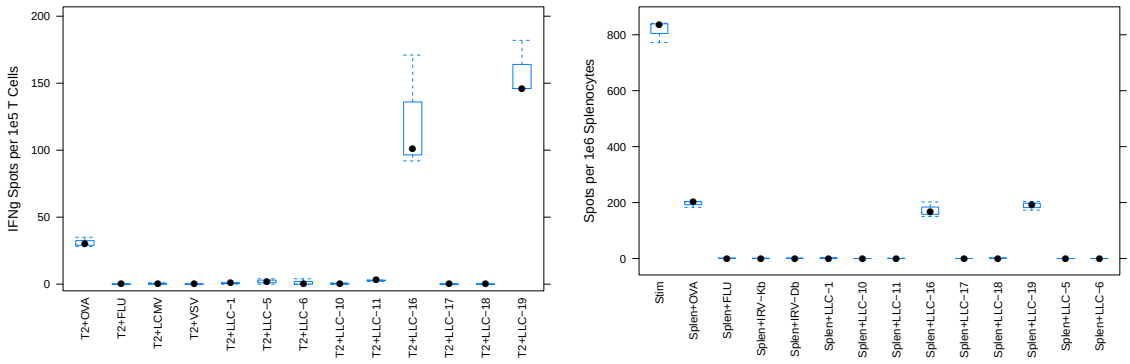

Figure 3.2: IFN-γ ELISpot results from the LLC candidate fusions, using T2 cells as antigen presenting cells (left) and splenocytes as a source of APC and effector cells (right)

To determine if these peptides were capable of endogenous presentation, the ELISpot assay was repeated using whole splenocytes from vaccinated mice to provide both the APCs and effector cells ($CD8^+$ and $CD4^+$ T cells). Interestingly, both LLC-16 and LLC-19 induced strong IFN-γ responses when pulsed onto whole splenocytes, however LLC-6 and 10 were undetectable (Figure 3.2). These peptides induced only moderate responses in experiments using T2-K^b as antigen presenting cells, so it is possible that the splenocyte assay was not sensitive enough to capture these responses.

Eight fusion candidates were identified in the MC38 cell line, which were tested using the same methods. ELISpot results, when using T2-K^b and T2-D^b as APCs, identified MC38-10 and MC38-13 immunogenic peptides (Figure 3.3), presented by T2-K^b . Additionally, MC38-1 induced a small but reproducible IFN-γ response when presented on T2-D^b . Replication of this experiment confirmed the immunogenicity

of these peptides, while also identifying MC38-7 as potentially immunogenic (data not shown).

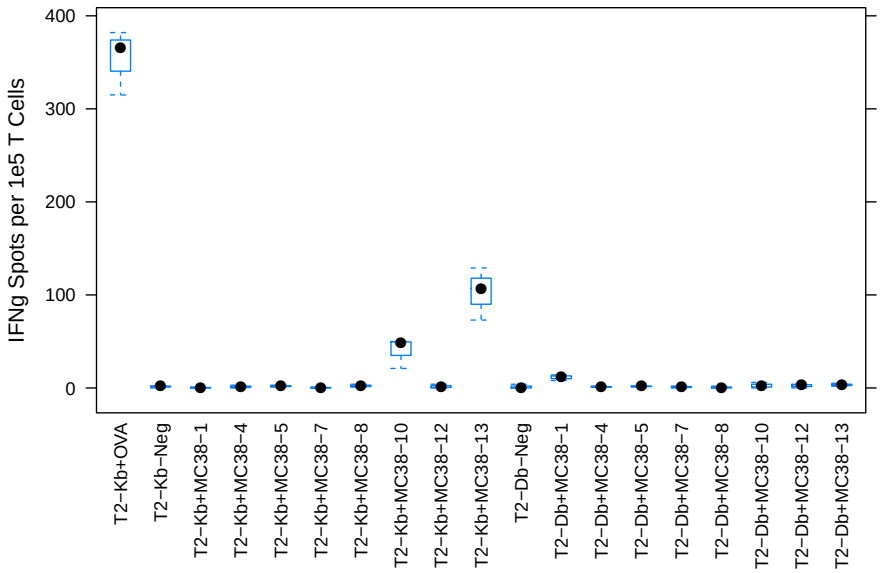

Figure 3.3: IFN-γ ELISpot results from the MC38 candidate fusions, using T2-D^b and T2-K^b as antigen presenting cells

Of the 4 LLC immunogenic hits, only LLC-10 was predicted to bind H-2-K^b, Interestingly, two peptides contained in the LLC-10 20mer were predicted to bind, TAPDYFLL (predicted affinity 193nM) and the longer TAPDYFLLQV (349nM). The LLC-18 peptide (YGAGYHMTL) was predicted to bind K^b with 284nM affinity, however, it did not induce a response in IFN-γ ELISpot experiments.

Three LLC derived peptides were immunogenic in IFN-γ ELISpot assays, despite poor binding predictions from `NetMHC`. The highest affinity peptide predicted from LLC-6 was IKPEWLVKI, with a predicted affinity of 3923nM. LLC-16 and 19 were even weaker binders, with best predicted affinities of 5325 and 14820nM, respectively.

MC38-derived peptides MC38-10 (GTFDRSVTL) and 13 (AVMGRTLSM) were both predicted to bind to H-2-K^b (477nM and 325nM, respectively). Both of these peptides induced an IFN-γ response in ELISpot assays.

MC38-7, which was identified as immunogenic in ELISpot assays, had a weak predicted affinity of 4837nM to K^b (peptide QSLSCCVCGGL). Interestingly, MC38-1 was identified as immunogenic when presented by T2-D^b , yet its best predicted binder was IVVVGTGT, with a predicted affinity of 31694nM. This peptide was actually predicted to bind to K^b with a higher affinity (7279nM), but no evidence was seen for this in IFN-γ ELIspots.

3.3.3 Assessment of Anti-tumor Activity

To assess the therapeutic potential of the fusion peptides identified, we performed a preventative vaccine experiment in which all candidate fusions (regardless of immunogenicity) for each tumor were administered in two doses (prime and boost) before tumor challenge. A vaccine containing the Ova peptide and adjuvant was used as a negative control.

No difference in tumor size was observed between fusion peptide groups and the negative control in either the LLC or MC38 models (Figure 3.4) when tumors were followed for 16 days after injection.

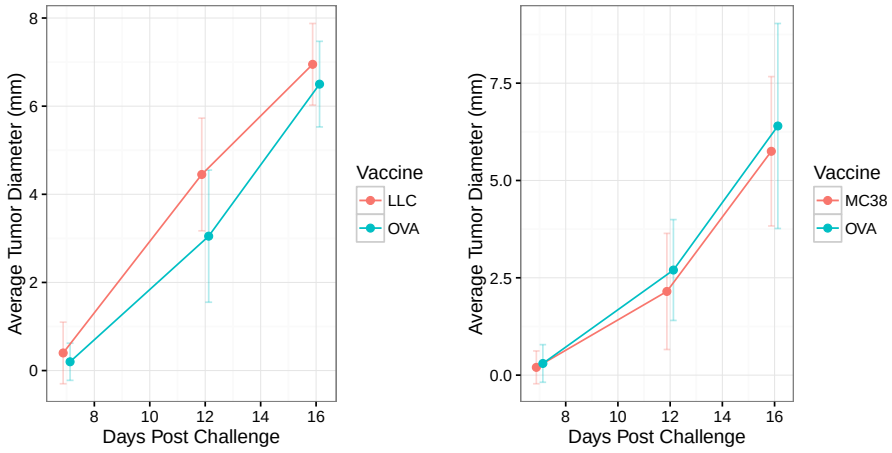

Figure 3.4: Average tumor diameter of LLC tumors (left) or MC38 tumors (right) when administered the fusion vaccine (red) or a control vaccine (blue). Error bars represent mean ± SD.

To validate the vaccine administration method and timing, the preventative vaccine experiment was repeated with the addition of a positive control vaccine group. This vaccine included peptides derived from point mutations identified in MC38, which were shown to induce an anti-tumor immune response. The MC38 and LLC fusion vaccines again demonstrated no anti-tumor effect, however the positive control group significantly reduced tumor growth (Figure 3.5).

To assess the possibility that the presence of tumor was suppressing the immune

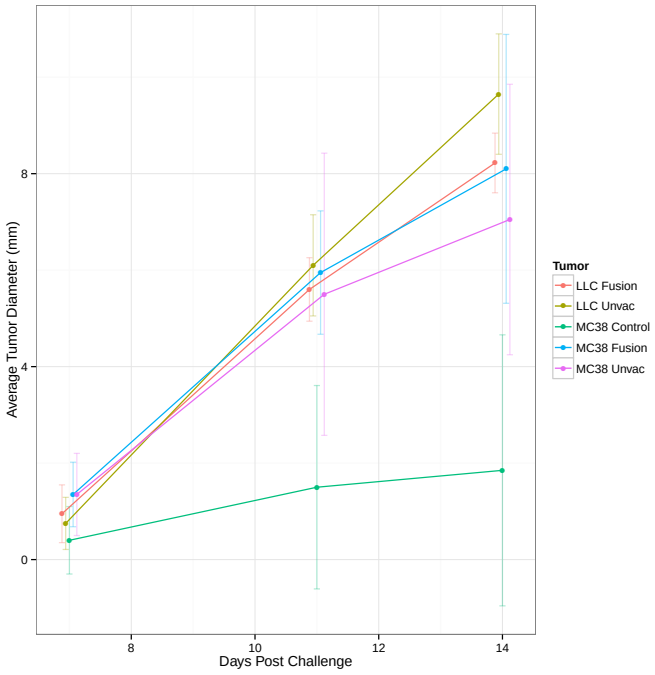

Figure 3.5: Average tumor diameters of mice treated with fusion vaccines or control vaccines. Error bars represent mean ± SD.

response in these mice, we confirmed that the immune response was still present 2 weeks after tumor challenge (3 weeks after the final vaccine dose) by performing an IFN-γ ELISpot. Because the immune response is expected to be lower after 3 weeks, an unchallenged cohort was included in this experiment.

In the LLC group, strong responses were seen against LLC-10, 16 and 19 at the pre-challenged time point. These responses diminished 2 weeks later, however no difference was seen between tumor challenged and unchallenged mice (Figure 3.6, top panel, bottom row). A similar pattern was observed in mice treated with a control vaccine (Figure 3.6, top panel, top row).

In the MC38 group, little reduction was seen in the number of IFN-γ spots at the later time point in unchallenged mice (Figure 3.6, bottom panel, bottom row). Tumor challenged mice exhibited a small reduction in the response, consistent with a state of immune suppression. Interestingly, the control vaccine group exhibited a

much larger drop in response at the later time point (Figure 3.6, bottom panel, top row).

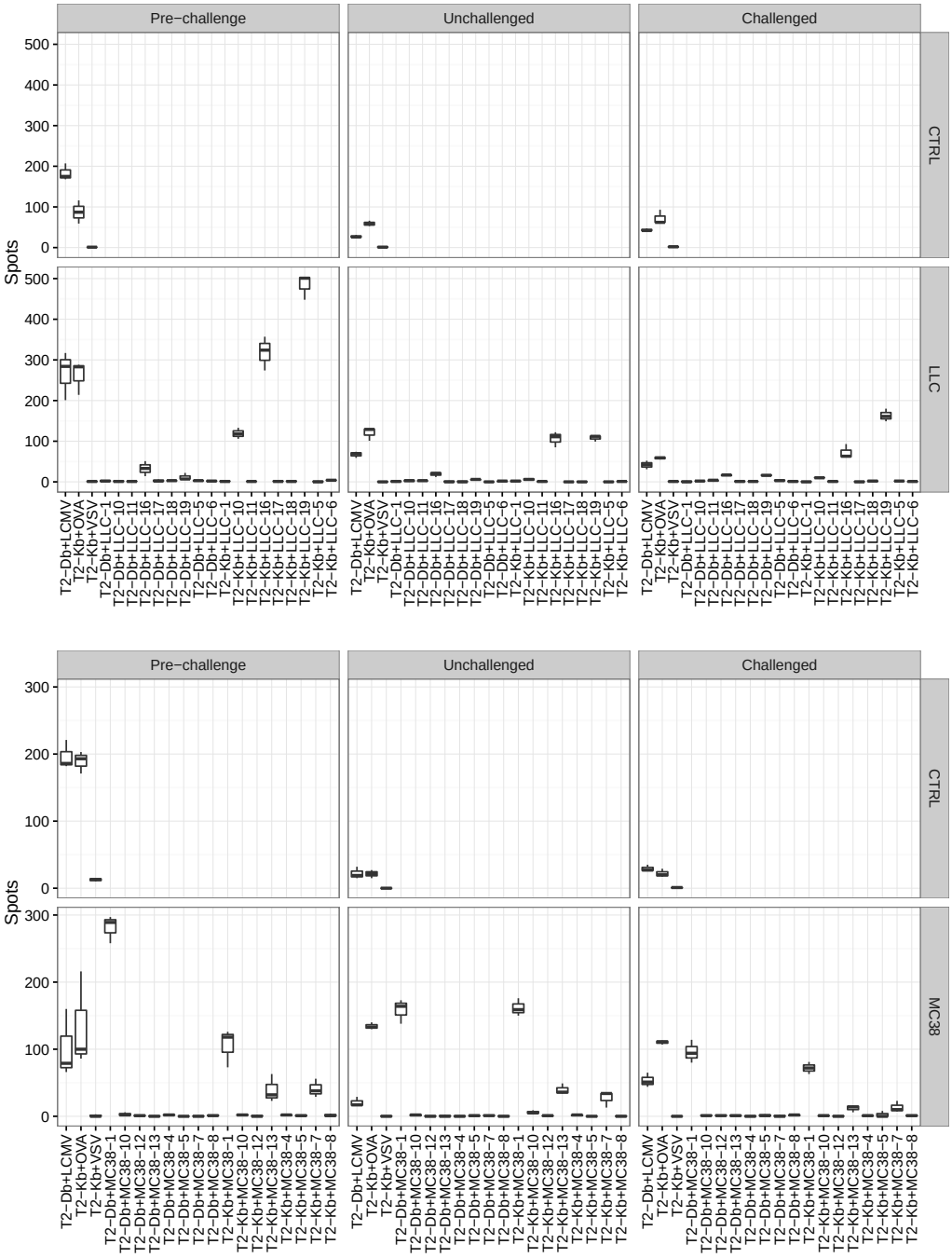

Figure 3.6: IFN-γ ELISpot of vaccine treated mice before tumor challenge (Pre-challenge), after 2 weeks without challenge (Unchallenged) or 2 weeks after tumor challenge (Challenged). Top rows represent mice given a control vaccine, bottom rows represent mice given the respective fusion vaccines.

3.3.4 Validation of Fusions using Whole Genome Sequencing

The lack of anti-tumor response from the fusion peptide vaccines could be caused by candidate peptides not being expressed, processed or presented on MHC molecules. To determine if the peptides identified in Section 3.3.1 using `tophat-fusion` were truly present in the cell lines, we performed whole genome sequencing (WGS) on genomic DNA from MC38 and LLC.

We identified putative fusion in the WGS data by identifying mate pairs which aligned either to different chromosomes or to distant sites of the same chromosome. Using permissive filters, we identified 1018 possible fusions in LLC and 262 in MC38.

Of the 1018 putative fusions from the WGS of the LLC cell line, only 2 overlapped with the list of candidate fusions derived from RNAseq data in Section 3.3.1. LLC-11 was supported by 17 read pairs in the WGS data, and LLC-13 was supported by only 4 pairs (LLC-13 was never tested *in vivo* because no translation was possible from this fusion event). LLC-11 showed no immunological activity in IFN-γ ELISpot assays.

Of the 262 candidates from WGS of the MC38 cell line, only 1 overlapped with RNAseq hits. MC38-8 was supported by 12 read pairs in the WGS data, but showed no biological activity in immune assays.

3.4 Discussion

Sequencing the transcriptome of the LLC and MC38 cell lines identified 17 putative fusions which were targeted in a preventative cancer vaccine model. Eight of those peptides induced a measurable immune response when administered with an adjuvant, however, none of these candidates were able to protect mice from challenge with their respective tumors. Immunization with previously validated peptides, known to be expressed and presented by the MC38 cell line, demonstrated that the model used is capable of detecting protective peptides. Finally, attempts to validate the candidate fusions using WGS data failed to confirm the validity of any of the 8 immunogenic peptides.

Algorithms like `tophat-fusion` are designed primarily for sensitivity (their ability to detect fusions, in spite of the background noise associated with RNAseq data), rather than specificity (the proportion of the candidates that are valid). In the context of peptide vaccines, some lack of specificity is acceptable, as the inclusion of irrelevant peptides (from invalid fusions identified in the data) does not, in theory, pose a risk in terms of the safety or efficacy of the vaccine. The sensitivity of the algorithm in our study is impossible to determine, as the true number of genomic rearrangements is not known. The specificity can be estimated by examining the overlap between hits obtained from RNAseq data and hits obtained from WGS. Unfortunately, this estimate indicates a very low specificity, as a shockingly small number of RNAseq hits had any evidence for validity in the WGS data. Because so few of the hits were valid, the sensitivity was unacceptably high.

The experiment did succeed in demonstrating an efficient method for identifying immunogenic fusions. Control vaccines also confirmed that immunogenic peptides, in combination with adjuvant can reduce tumor growth in these preclinical models. These *in vivo* and *in vitro* models are capable of identifying immunogenic and effective peptides, but must be provided with improved peptide candidates to succeed.

Chapter 4

Summary and Conclusions

Immunotherapy has shown great promise in a number of different cancer types, but its benefits are often only seen by a subset of treated patients. This heterogeneity of responses results from the fact that every patient has a truly unique immune system, developed over a lifetime of antigen exposure. The two areas of research described here represent parallel attempts to understand and overcome this problem. Profiling of immune repertoires allows better understanding of what constitutes a healthy immune system, while also predicting the likelihood of response to immunotherapy. Personalized vaccines promise to deliver responses to more patients by assuring that cases of rare mutations are not disadvantaged.

In Chapter 2, a large cohort of pancreatic cancer immunotherapy patients were subjected to TCR repertoire profiling before and after treatment. The results of this analysis have helped to build understanding about what effects are induced by different forms of immune modulation. Anti-CTLA4 treatment appears to have the largest effect on the T-cell compartment, while vaccination efforts, chemotherapy and radiation are all much harder to measure. Efforts to extract meaningful data from tumor infiltrating T-cell repertoires were hampered by low sample sizes, but statistical evaluation of this data indicates that the quality of data from FFPE is no worse than

that of peripheral cells, indicating that similar effect sizes are detectable from both cell sources. The power analysis provided in Chapter 2.4 should be extremely useful for the development of future experiments by helping to decide the number of samples necessary to identify a desired effect size.

A model was developed to predict clinical response, and efforts to test this model on new validation datasets will determine both its success in pancreatic cancer and its applicability in other contexts. The model is built upon summary metrics of TCR data, which only scratch the surface of available information created by this technology. HTTCS will find broader uses in cancer immunotherapy, especially in cases where treatment involves the administration of known antigens (peptide vaccination or similar) and cases of adoptive transfer of lymphocytes, where the input population can be tracked.

Analysis of TCR data in the literature has been somewhat scattered in its approach and methodology. Early reports conflated the concepts of Richness (the number of unique TCRs identified) and diversity (which takes into account both Richness and the distribution of clones within it). Richness was excluded from this analysis early on because it was heavily correlated with the number of input cells sequenced. The `immunoSeqR` package attempts to provide a framework for analyzing TCR repertoire data by combining the summary metrics and metadata aware functions to compare and plot them.

Chapter 3 described attempts to develop a "personalized" vaccine targeting the fusion proteins that arise as the result of genomic rearrangements in cancer. An initial RNAseq approach to identify candidate peptides failed to provide a significant number of valid hits, rendering downstream efforts to characterize them unsuccessful. The nature of modern sequencing technology amplifies the difficulty of detecting genomic rearrangements with high sensitivity and specificity. Current technologies rely on sheared DNA libraries, which can confuse algorithms designed to detect natural

fusions. New methods of library preparation, improvements to existing computational methods, and use of newer sequencing chemistries capable of longer continuous reads may help ameliorate the issue described above.

Both arms of this study have furthered our understanding of some of the largest issues confronting cancer immunotherapy research. Continued efforts to develop sequencing based biomarkers using HTTCS will undoubtedly find successful niches. Improved antigen identification efforts will facilitate the development of personalized cancer vaccines to bridge the gap between responders and non-responders. Both efforts will benefit greatly from research at the interface of biology and technology and hopefully move forward into clinical settings to improve outcomes for cancer patients.

References

[1] N Howlader et al. *SEER Cancer Statistics Review, 1975-2013*. Sept. 2016.

[2] Audrey Vincent et al. "Pancreatic Cancer". In: *The Lancet* 378.9791 (2011), pp. 607–620.

[3] G. Bond-Smith et al. "Pancreatic Adenocarcinoma". In: *BMJ* 344.may16 1 (May 2012), e2476–e2476.

[4] Matthew F. Krummel and James P. Allison. "CD28 and CTLA-4 Have Opposing Effects on the Response of T Cells to Stimulation". In: *Journal of Experimental Medicine* 182.2 (1995), pp. 459–466.

[5] D. R. Leach, M. F. Krummel, and J. P. Allison. "Enhancement of Antitumor Immunity by CTLA-4 Blockade". In: *Science (New York, N.Y.)* 271.5256 (Mar. 1996), pp. 1734–1736.

[6] Jennifer Couzin-Frankel. *Cancer Immunotherapy*. American Association for the Advancement of Science, 2013.

[7] E. M. Jaffee et al. "Novel Allogeneic Granulocyte-Macrophage Colony-Stimulating Factor-Secreting Tumor Vaccine for Pancreatic Cancer: A Phase I Trial of Safety and Immune Activation". In: *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 19.1 (Jan. 2001), pp. 145–156.

[8] E. R. Lutz et al. “A Lethally Irradiated Allogeneic Granulocyte-Macrophage Colony Stimulating Factor-Secreting Tumor Vaccine for Pancreatic Adenocarcinoma: A Phase II Trial of Safety, Efficacy, and Immune Activation”. In: *Annals of Surgery* 253.2 (Feb. 2011), pp. 328–335.

[9] Richard E. Royal et al. “Phase 2 Trial of Single Agent Ipilimumab (Anti-CTLA-4) for Locally Advanced or Metastatic Pancreatic Adenocarcinoma”. In: *Journal of immunotherapy* 33.8 (2010), pp. 828–833.

[10] Dung T. Le et al. “Evaluation of Ipilimumab in Combination With Allogeneic Pancreatic Tumor Cells Transfected With a GM-CSF Gene in Previously Treated Pancreatic Cancer:” in: *Journal of Immunotherapy* 36.7 (Sept. 2013), pp. 382–389.

[11] Kevin C. Soares et al. “PD-1/PD-L1 Blockade Together With Vaccine Therapy Facilitates Effector T-Cell Infiltration Into Pancreatic Tumors:” in: *Journal of Immunotherapy* 38.1 (Jan. 2015), pp. 1–11.

[12] D. T. Le et al. “Safety and Survival With GVAX Pancreas Prime and Listeria Monocytogenes-Expressing Mesothelin (CRS-207) Boost Vaccines for Metastatic Pancreatic Cancer”. In: *Journal of Clinical Oncology* 33.12 (Apr. 2015), pp. 1325–1333.

[13] B. Zink et al. “A Direct Estimate of the Human O4, T CeLL Receptor Diversity”. In: (1999).

[14] Harlan S. Robins et al. “Comprehensive Assessment of T-Cell Receptor Beta-Chain Diversity in Alphabeta T Cells”. In: *Blood* 114.19 (Nov. 2009), pp. 4099–4107.

[15] D. Wu et al. “Detection of Minimal Residual Disease in B Lymphoblastic Leukemia by High-Throughput Sequencing of IGH”. In: *Clinical Cancer Research* 20.17 (Sept. 2014), pp. 4540–4548.

[16] William S. DeWitt et al. “Dynamics of the Cytotoxic T Cell Response to a Model of Acute Viral Infection”. In: *Journal of Virology* 89.8 (Apr. 2015). Ed. by R. M. Sandri-Goldin, pp. 4517–4526.

[17] Harlan Robins. “Immunosequencing: Applications of Immune Repertoire Deep Sequencing”. In: *Current opinion in immunology* 25.5 (Oct. 2013), pp. 646–652.

[18] L. Robert et al. “CTLA4 Blockade Broadens the Peripheral T-Cell Receptor Repertoire”. In: *Clinical Cancer Research* 20.9 (May 2014), pp. 2424–2432.

[19] E. Cha et al. “Improved Survival with T Cell Clonotype Stability After Anti-CTLA-4 Treatment in Cancer Patients”. In: *Science Translational Medicine* 6.238 (May 2014), 238ra70–238ra70.

[20] Manuarii Manuel et al. “Lymphopenia Combined with Low TCR Diversity (Divpenia) Predicts Poor Overall Survival in Metastatic Breast Cancer Patients”. In: *OncoImmunology* 1.4 (July 2012), pp. 432–440.

[21] Michael A. Postow et al. “Peripheral T Cell Receptor Diversity Is Associated with Clinical Outcomes Following Ipilimumab Treatment in Metastatic Melanoma”. In: *Journal for Immunotherapy of Cancer* 3 (2015), p. 23.

[22] Lidia Robert et al. “Distinct Immunological Mechanisms of CTLA-4 and PD-1 Blockade Revealed by Analyzing TCR Usage in Blood Lymphocytes”. In: *OncoImmunology* 3.6 (June 2014), e29244.

[23] E. R. Lutz et al. “Immunotherapy Converts Nonimmunogenic Pancreatic Tumors into Immunogenic Foci of Immune Regulation”. In: *Cancer Immunology Research* (June 2014).

[24] Ryan O Emerson et al. “High-Throughput Sequencing of T-Cell Receptors Reveals a Homogeneous Repertoire of Tumour-Infiltrating Lymphocytes in Ovarian Cancer: Tumour-Restricted and Homogeneous TILs in Ovarian Cancer”. In: *The Journal of Pathology* 231.4 (Dec. 2013), pp. 433–440.

[25] Marco Gerlinger et al. “Ultra-Deep T Cell Receptor Sequencing Reveals the Complexity and Intratumour Heterogeneity of T Cell Clones in Renal Cell Carcinomas: Ultra-Deep Sequencing of T Cell Repertoires in Renal Cancer”. In: *The Journal of Pathology* 231.4 (Dec. 2013), pp. 424–432.

[26] Paul C. Tumeh et al. “PD-1 Blockade Induces Responses by Inhibiting Adaptive Immune Resistance”. In: *Nature* 515.7528 (Nov. 2014), pp. 568–571.

[27] C. E. Shannon. “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27.3 (July 1948), pp. 379–423.

[28] Yosef Hochberg and Yoav Benjamini. “More Powerful Procedures for Multiple Significance Testing”. In: *Statistics in medicine* 9.7 (1990), pp. 811–818.

[29] Dung T. Le and Elizabeth M. Jaffee. “Harnessing Immune Responses in the Tumor Microenvironment: All Signals Needed”. In: *Clinical Cancer Research* 19.22 (Nov. 2013), pp. 6061–6063.

[30] Wolfgang Walther, ed. *Current Strategies in Cancer Gene Therapy*. Vol. 209. Recent Results in Cancer Research. Cham: Springer International Publishing, 2016.

[31] Hiep Khong and Willem W. Overwijk. “Adjuvants for Peptide-Based Cancer Vaccines”. In: *Journal for ImmunoTherapy of Cancer* 4.1 (Dec. 2016).

[32] J. C. Castle et al. “Exploiting the Mutanome for Tumor Vaccination”. In: *Cancer Research* 72.5 (Mar. 2012), pp. 1081–1091.

[33] Madiha Derouazi et al. “Novel Cell-Penetrating Peptide-Based Vaccine Induces Robust CD4+ and CD8+ T Cell-Mediated Antitumor Immunity”. In: *Cancer Research* (June 2015).

[34] Dan P. Zandberg et al. “A Phase I Dose Escalation Trial of MAGE-A3- and HPV16-Specific Peptide Immunomodulatory Vaccines in Patients with Recurrent/Metastatic (RM) Squamous Cell Carcinoma of the Head and Neck (SCCHN)”. In: *Cancer Immunology, Immunotherapy* 64.3 (Mar. 2015), pp. 367–379.

[35] Ryuji Takahashi et al. “Feasibility Study of Personalized Peptide Vaccination for Metastatic Recurrent Triple-Negative Breast Cancer Patients”. In: *Breast Cancer Research* 16.4 (2014), R70.

[36] S. Kibe et al. “Phase II Study of Personalized Peptide Vaccination for Previously Treated Advanced Colorectal Cancer”. In: *Cancer Immunology Research* 2.12 (Dec. 2014), pp. 1154–1162.

[37] J. H. Sampson et al. “Immunologic Escape After Prolonged Progression-Free Survival With Epidermal Growth Factor Receptor Variant III Peptide Vaccination in Patients With Newly Diagnosed Glioblastoma”. In: *Journal of Clinical Oncology* 28.31 (Nov. 2010), pp. 4722–4729.

[38] M. K. Gjertsen et al. “Intradermal Ras Peptide Vaccination with Granulocyte-Macrophage Colony-Stimulating Factor as Adjuvant: Clinical and Immunological Responses in Patients with Pancreatic Adenocarcinoma”. In: *International Journal of Cancer. Journal International Du Cancer* 92.3 (May 2001), pp. 441–450.

[39] Monica Bocchia et al. “Complete Molecular Response in CML after P210 BCR–ABL1-Derived Peptide Vaccination”. In: *Nature Reviews Clinical Oncology* 7.10 (Oct. 2010), pp. 600–603.

[40] Martin KH Maus et al. “Identification of Novel Variant of EML4-ALK Fusion Gene in NSCLC: Potential Benefits of the RT-PCR Method”. In: *International journal of biomedical science: IJBS* 8.1 (2012), p. 1.

[41] W. A. Freed-Pastor and C. Prives. “Mutant P53: One Name, Many Proteins”. In: *Genes & Development* 26.12 (June 2012), pp. 1268–1286.

[42] H Rammensee et al. “SYFPEITHI: Database for MHC Ligands and Peptide Motifs”. In: *Immunogenetics* 50.3-4 (Nov. 1999), pp. 213–219.

[43] Claus Lundegaard et al. “Major Histocompatibility Complex Class I Binding Predictions as a Tool in Epitope Discovery: MHC Class I Binding Predictions”. In: *Immunology* 130.3 (July 2010), pp. 309–318.

[44] Morten Nielsen et al. “The Role of the Proteasome in Generating Cytotoxic T-Cell Epitopes: Insights Obtained from Improved Predictions of Proteasomal Cleavage”. In: *Immunogenetics* 57.1-2 (Apr. 2005), pp. 33–41.

[45] Morten Nielsen et al. “Reliable Prediction of T-Cell Epitopes Using Neural Networks with Novel Sequence Representations”. In: *Protein Science: A Publication of the Protein Society* 12.5 (May 2003), pp. 1007–1017.

[46] S. Rosenberg, P Spiess, and R Lafreniere. “A New Approach to the Adoptive Immunotherapy of Cancer with Tumor-Infiltrating Lymphocytes”. In: *Science* 233.4770 (Sept. 1986), pp. 1318–1321.

[47] John S. Bertram and Przemyslaw Janik. “Establishment of a Cloned Line of Lewis Lung Carcinoma Cells Adapted to Cell Culture”. In: *Cancer Letters* 11.1 (Nov. 1980), pp. 63–73.

[48] W. James Kent et al. “The Human Genome Browser at UCSC”. In: *Genome research* 12.6 (2002), pp. 996–1006.

[49] Ben Langmead et al. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome”. In: *Genome biology* 10.3 (2009), R25.

[50] C. Trapnell, L. Pachter, and S. L. Salzberg. “TopHat: Discovering Splice Junctions with RNA-Seq”. In: *Bioinformatics* 25.9 (May 2009), pp. 1105–1111.

[51] James T Robinson et al. “Integrative Genomics Viewer”. In: *Nature Biotechnology* 29.1 (Jan. 2011), pp. 24–26.

[52] S Buus et al. “Sensitive Quantitative Predictions of Peptide-MHC Binding by a ’Query by Committee’ Artificial Neural Network Approach”. In: *Tissue antigens* 62.5 (Nov. 2003), pp. 378–384.

[53] Mahesh Yadav et al. “Predicting Immunogenic Tumour Mutations by Combining Mass Spectrometry and Exome Sequencing”. In: *Nature* 515.7528 (Nov. 2014), pp. 572–576.

Alexander Hopkins

email: ahopkins@protonmail.com phone: (269)501-3602

Education

The Johns Hopkins University School of Medicine 9/2013 - 8/2017 (Expected)
Candidate for Ph.D. in Cellular and Molecular Medicine

Massachusetts Institute of Technology 9/2006 - 6/2010
Bachelor of Science in Biology

Research Experience

The Johns Hopkins University School of Medicine PhD Candidate
Laboratory of Elizabeth Jaffee, MD 9/2013 - Present

- Analyzed high throughput T-cell receptor β sequencing data to understand the mechanisms of successful anti-tumor vaccination in pancreatic cancer
  - Developed an `R` package to compare samples in TCR sequencing experiments
  - Identified potential new biomarkers for pancreatic cancer immunotherapy
- Developed a novel preclinical cancer vaccine targeting immunogenic gene fusions
  - Identified unique gene fusions from RNAseq and Whole Genome Sequencing data
  - Developed and tested fusion vaccines in multiple murine cancer models

The Broad Institute of MIT and Harvard Research Associate I & II
Laboratory of Todd Golub, MD 9/2010 - 6/2013

- Worked in support of the Broad's Cancer Cachexia and Hepatotoxicity projects in the laboratory and as a computational analyst
- Assisted in development of methods for high throughput culture of human adipocytes, myocytes and hepatocytes
- Performed genetic screen of secreted proteins in cancer cell lines, generated more than 9000 gene expression profiles and developed software to compare perturbed states to known gene expression signatures, implicating a handful of putative cachexia-causing secreted proteins
- Developed software to analyze in-cell western blot data, including a graphical user interface to increase usability for laboratory scientists
- Promoted to Research Associate II, November 2012

MIT Department of Biology Student Researcher
Laboratory of Michael Laub, PhD 1/2009 - 9/2010

- One of two winners of the John Asinari award for outstanding research in the field of life sciences, 2010
- Worked with a graduate student to uncover the mechanisms of DNA damage in the model organism *Caulobacter crescentus*
- Performed a large plasmid based synthetic lethal screen, discovered the function of a novel gene, named *sid*A which causes a cell cycle arrest via a previously undescribed mechanism

Software Development

`immunoSeqR` - An `R` package for analyzing high throughput T-cell receptor β sequencing data, including metrics, metadata-aware plotting and single document reports. Available at `https://github.com/ahopki14/immunoSeqR`

`Hexis` - A `MATLAB` GUI for comparing thousands of gene expression signatures to other known signatures

`icw` - A `MATLAB` GUI for analyzing Licor in-cell western data

Publications and Presentations

- *T Cell Receptor Profiling in Pancreatic Cancer Patients Receiving Immune Checkpoint Inhibition Combination Therapies.* **Alexander Hopkins**, Mark Yarchoan, Jennifer N Uram, Dung Le, Elizabeth M Jaffee and Eric R Lutz. (Manuscript in preparation)
- *Relationship between lymphopenia and objective response rate with programmed death-1 (PD-1) inhibitor therapy: A single-center retrospective analysis.* Mark Yarchoan, Adam Diehl, Burles Avner Johnson III, Blake Scott, **Alexander Hopkins**, Nilofer Saba Azad, Elizabeth M. Jaffee, Stuart A. Grossman. American Society of Clinical Oncology Abstract, 2017.
- *Relationship of lymphocyte and eosinophil counts and immune-related adverse events in recipients of programmed death-1 (PD-1) inhibitor therapy: A single-center retrospective analysis.* Adam Diehl, Mark Yarchoan, Ting Yang, Blake Scott, Burles Avner Johnson III, **Alexander Hopkins**, Nilofer Saba Azad, Elizabeth M. Jaffee, Stuart A. Grossman. American Society of Clinical Oncology Abstract, 2017.
- *Molecular Mediators of Cancer Cachexia.* David Thomas, **Alexander Hopkins**, Nemanja Marjanovic, Todd Golub. (Manuscript in preparation)
- *Predicting Drug Toxicity using Multiplexed Gene Expression and High Content Imaging in Primary Human Hepatocytes.* David Thomas, **Alexander Hopkins**, Nemanja Marjanovic, Todd Golub. (Manuscript in preparation)
- *Designing Diverse, High Content Assays to Uncover the Molecular Mediator of Cancer Cachexia.* **Alexander Hopkins**, David Rodriguez-Fuentes, David Thomas, Todd Golub. Poster presented at the Broad Institute Scientific Retreat 2012
- Joshua Modell, **Alexander Hopkins**, Michael Laub. *A DNA damage checkpoint in Caulobacter crescentus inhibits cell division through a direct interaction with FtsW.* Genes Dev. 2011 Jun 15;25(12):1328-43.

Computer Skills

- Experienced in analysis of data from T-cell receptor sequencing, gene expression, whole genome sequencing, qPCR and survival experiments
- Skilled at using Linux, Mac OS, and Windows operating systems. Experienced with `R`, `MATLAB` and `Bash` languages, as well as `R Markdown`, `Sweave` and LaTeX document preparation systems
- Webmaster for Cellular and Molecular Medicine Graduate Program, Johns Hopkins University School of Medicine, 2013-Present

Laboratory Skills

- Over seven years experience in mammalian cell culture, including human primary cells
- Experienced with a variety of model organisms, including bacteria, *C. elegans* and mice
- Familiar with a variety of molecular biology techniques

References

Elizabeth Jaffee, MD
The Dana and Albert "Cubby" Broccoli Professor of Oncology
Co-Director of the Gastrointestinal Cancers Program
Associate Director for Translational Research
The Sidney Kimmel Cancer Center at Johns Hopkins

Eric Lutz, PhD
Senior Principal Scientist
WindMIL Therapeutics

Todd Golub, MD
Chief Scientific Officer
Director of the Cancer Program
The Broad Institute of MIT and Harvard